

Box Sep. A

"EXPRESS MAIL CERTIFICATE"

"EXPRESS MAIL" MAILING LABEL NUMBER EH856102078US

DATE OF DEPOSIT FEBRUARY 9, 1998

**I HEREBY CERTIFY THAT THIS PAPER OR FEE IS BEING DEPOSITED
WITH THE UNITED STATES POSTAL SERVICE "EXPRESS MAIL POST
OFFICE TO ADDRESSEE" SERVICE UNDER 37 CFR 1.10 ON THE DATE
INDICATED ABOVE AND IS ADDRESSED TO THE ASSISTANT COMMISSIONER
FOR PATENTS, BOX PATENT APPLICATION, WASHINGTON, DC 20231.**

**NAME OF PERSON MAILING PAPER OR FEE
(TYPE OR PRINT) TRACY WESTPHAL**

SIGNATURE

Tracy Westphal

Attorney Docket No. 0815

February 9, 1998

Assistant Commissioner for Patents
Box Patent Application
Washington, DC 20231

UTILITY PATENT APPLICATION TRANSMITTAL

Inventor(s): Jung *et al.*

Title: ALTERATION OF AMINO ACID COMPOSITIONS IN SEEDS

APPLICATION ELEMENTS

1. ☒ Fee Transmittal Form (Submit an original, and a duplicate for fee processing)
2. ☒ Specification [Total Pages 59]
(Preferred arrangement set forth below)
 - Descriptive title of the Invention
 - Cross Reference to Related Applications
 - Statement Regarding Fed sponsored R & D
 - Reference to Microfiche Appendix
 - Background of the Invention
 - Brief Summary of the Invention
 - Brief Description of the Drawings (if filed)
 - Detailed Description
 - Claim(s)
 - Abstract of the Disclosure
3. ☐ Drawing(s) (35 USC 113) [Total Sheets]
 - a. ☐ Formal
 - b. ☐ Informal

02/09/98
JCS74 U.S. PTO

09020316 020998

4. ☐ Oath or Declaration [Total Pages]
- a. ☐ Newly executed (original or copy)
- b. ☐ Copy from a prior application (37 CFR 1.63(d))
(for continuation/divisional with Box 17 completed)
[Note Box 5 below]
- i. ☐ DELETION OF INVENTOR(S)
Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).
5. ☐ Incorporation By Reference (useable if Box 4B is checked)
The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.
6. ☐ Microfiche Computer Program (Appendix)
7. ☒ Nucleotide and/or Amino Acid Sequence Submission (if applicable, all necessary)
- a. ☒ Computer Readable Copy
- b. ☒ Paper Copy (identical to computer copy)
- c. ☒ Statement verifying identity to above copies

ACCOMPANYING APPLICATION PARTS

8. ☐ Assignment Papers (cover sheet & document(s))
9. ☐ 37 CFR 3.73(b) Statement ☐ Power of Attorney
(where there is an assignee)
10. ☐ English Translation Document (if applicable)
11. ☒ Information Disclosure Statement (IDS/PTO-1449) ☒ Copies of IDS Citations
12. ☐ Preliminary Amendment
13. ☒ Return Receipt Postcard (MPEP 503) (Should be specifically itemized)
14. ☐ Small Entity Statement(s) ☐ Statement filed in prior application
Status still proper and desired

Attorney Docket No.: 0815

15. ☐ Certified Copy of Priority document(s)

16. ☐ Other:

17. If a **CONTINUING APPLICATION**, check the appropriate box and supply the requisite information: ☐ Continuation ☐ Divisional ☐ Continuation-in-part (CIP) of prior application No: _____.

18. CORRESPONDENCE ADDRESS

☐ Customer Number or Bar Code Label: _____
(Insert Customer No. or Attach bar code label here)
or

☒ Correspondence address below

Name: Marianne H. Michel

Address: Pioneer Hi-Bred International, Inc.
Corporate Intellectual Property
7100 N.W. 62nd Avenue
P.O. Box 1000

City Johnston State: Iowa Zip Code: 50131-1000

Country: USA Phone: (515) 334-4467 Facsimile: (515) 334-6883

19. FEE CALCULATION

☒ The total fee is calculated as shown below:

Basic Filing Fee					\$ 790.00
Total Claims	21 - 20	= 1	x	\$ 22.00	22.00
Independent Claims	4 - 3	= 1	x	\$ 82.00	82.00

☐ Multiple Dependent Claims present + \$270.00 \$

TOTAL FILING FEE \$ 894.00

☒ Charge \$894.00 to Deposit Account No. 16-1852.

Attorney Docket No.: 0815

- ☒ The Commissioner is hereby authorized to charge any additional fees under 37 CFR 1.16 or 1.17 which may be required by this paper, or credit any overpayment, to our Deposit Account No. 16-1852.
- ☒ A duplicate of this transmittal is enclosed

Respectfully submitted, .

Marianne H. Michel

Marianne H. Michel
Attorney for Applicant(s)
Registration No. 35,286

PIONEER HI-BRED INTERNATIONAL, INC.
Corporate Intellectual Property
7100 N.W. 62nd Avenue
P.O. Box 1000
Johnston, Iowa 50131-1000
Phone: (515) 334-4467
Facsimile: (515) 334-6883

ALTERATION OF AMINO ACID COMPOSITIONS IN SEEDS

Rudolf Jung
Larry R. Beach
Virginia M. Dress
A. Gururaj Rao
Jerome P. Ranch
David S. Ertl
Regina K. Higgins

BACKGROUND OF THE INVENTION

Feed formulations based on crop plants must typically be supplemented with specific amino acids to provide animals with essential nutrients which are necessary for their growth. This supplementation is necessary because, in general, crop plants contain low proportions of several amino acids which are essential for, and cannot be synthesized by, monogastric animals.

The seeds of crop plants contain different classes of seed proteins. The amino acid composition of these seeds reflects the composition of the prevalent classes of proteins. Amino acid limitations are usually due to amino acid deficiencies of these prevalent protein classes.

Among the amino acids necessary for animal nutrition, those that are of limited availability in crop plants include methionine, lysine, and threonine. Attempts to increase the levels of these amino acids by breeding, mutant selection, and/or changing the composition of the storage proteins accumulated in the seeds of crop plants, have met with limited success, or were accompanied by a loss in yield.

For example, although seeds of corn plants containing a mutant transcription factor, (opaque 2), or a mutant α -zein gene, (floury 2), exhibit elevated levels of total and bound lysine, there is an altered seed endosperm structure which is more susceptible to damage and pests. Significant yield losses are also typical.

An alternative means to enhance levels of free amino acids in a crop plant is the modification of amino acid biosynthesis in the plant. The introduction of a feedback-regulation-insensitive dihydrodipicolinic acid synthase ("DHDPS") gene, which encodes an enzyme that catalyzes the first reaction unique to the lysine biosynthetic pathway, into plants has resulted in an increase in the levels of free lysine in the leaves and seeds of those plants. An increase in the levels of free lysine in the embryo results in reduced amount of oil in the seed. Further free lysine can be lost during the wet milling process reducing the feed value of the gluten product of the process.

The expression of the *lysC* gene, which encodes a mutant bacterial aspartate kinase that is desensitized to feedback inhibition by lysine and threonine, from a seed-specific promoter in tobacco plants, has resulted in an increase in methionine and threonine biosynthesis in the seeds of those plants. See Karchi, *et al.*; The Plant J.; Vol. 3; p. 721; (1993). However, expression of the *lysC* gene results in only a 6-7% increase in the level of total threonine or methionine in the seed. The expression of the *lysC* gene in seeds has a minimal impact on the nutritional value of those seeds and, thus, supplementation of feed containing *lysC* transgenic seeds with amino acids, such as methionine and threonine, is still required.

There are additional molecular genetic strategies available for enhancing the amino acid quality of plant proteins. Each involves molecular manipulation of plant genes and the generation of transgenic plants.

Protein sequence modification involves the identification of a gene encoding a major protein, preferably a storage protein, as the target for modification to contain more codons of essential amino acids. An important aspect of this approach is to be

able to select a region of the protein that can be modified without affecting the overall structure, stability, function, and other cellular and nutritional properties of the protein.

The development of DNA synthesis technology allows the design and synthesis of a gene encoding a new protein with desirable essential amino acid compositions. For example, researchers have synthesized a 292-base pair DNA sequence encoding a polypeptide composed of 80% essential amino acids and used it with the nopaline synthetase (NOS) promoter to construct a chimeric gene. Expression of this gene in the tuber of transgenic potato has resulted in an accumulation of this protein at a level of 0.02% to 0.35% of the total plant protein. This low level accumulation is possibly due to the weak NOS promoter and/or the instability of the new protein.

Tobacco has been used as a test plant to demonstrate the feasibility of this approach by transferring a chimeric gene containing the bean phaseolin promoter and the cDNA of a sulfur-rich protein Brazil Nut Protein ("BNP"), (18 mol% methionine and 8 mol% cysteine) into tobacco. Amino acid analysis indicates that the methionine content in the transgenic seeds is enhanced by 30% over that of the untransformed seeds. This same chimeric gene has also been transferred into a commercial crop, canola, and similar levels of enhancement were achieved.

However, an adverse effect is that lysine content decreases. Additionally, BNP has been identified as a major food allergen. Thus it is neither practical nor desirable to use BNP to enhance the nutritional value of crop plants.

Thus, there is a need to improve the nutritional value of plant seeds. The genetic modification should not be accompanied by detrimental side effects such as allergenicity, anti-nutritional quality or poor yield.

SUMMARY OF THE INVENTION

5 It is an object of the present invention to provide a seed, the endosperm of which contains elevated levels of an essential amino acid.

It is a further object of the present invention to provide methods for increasing the nutritional value of feed.

10 It is a further object of the present invention to provide methods for genetically modifying seeds so as to increase amounts of essential amino acids which are present in relatively low amounts in unmodified seeds.

It is a further object of the present invention to provide methods for increasing the nutritional content of seeds without detrimental side effects such as allergenicity or anti-nutritional quality.

15 It is a further object of the present invention to provide methods for increasing the nutritional content of seeds while maintaining a high yield.

It is a further object of the present invention to provide a method for the expression of a polypeptide in a seed having levels of a preselected amino acid sufficient to reduce or obviate feed supplementation.

20 According to the present invention a transformed plant seed is provided, the endosperm of which is characterized as having an elevated level of at least one preselected amino acid compared to a seed from a corresponding plant which has not been transformed, wherein the amino acid is lysine, threonine, or tryptophan and optionally a sulfur-containing amino acid.

Also provided is a seed from a plant which has been transformed to express a heterologous protein in the endosperm of the seed, wherein the seed exhibits an elevated level of an essential amino acid.

An expression cassette is also provided comprising a seed endosperm-
5 preferred promoter operably linked to a structural gene encoding a polypeptide having an elevated level of a preselected amino acid. Transformed plants and seeds containing the expression cassette are also provided.

A method for elevating the level of a preselected amino acid in the endosperm of plant seed is also provided. The method comprises the
10 transformation of plant cells by introducing the expression cassette, recovering the transformed cells, regenerating a transformed plant and collecting the seeds therefrom.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, a "structural gene " means an exogenous or recombinant
15 DNA sequence or segment that encodes a polypeptide.

As used herein, "recombinant DNA" is a DNA sequence or segment that has been isolated from a cell, purified, synthesized or amplified.

As used herein, "isolated" means either physically isolated from the cell or synthesized *in vitro* on the basis of the sequence of an isolated DNA segment.

20 As used herein, the term "increased" or "elevated" levels of the preselected amino acid in a protein means that the protein contains an elevated amount of a preselected amino acid compared to the amount in an average protein.

As used herein, "increased" or "elevated" levels or amounts of preselected amino acids in a transformed plant or seed are levels which are greater than the levels or amounts in the corresponding untransformed plant or seed.

As used herein, "polypeptide" means proteins, protein fragments, modified proteins, amino acid sequences and synthetic amino acid sequences.

As used herein, "transformed plant" means a plant which comprises a structural gene which is introduced into the genome of the plant by transformation.

As used herein, "untransformed plant" refers to a wild type plant, i.e., one where the genome has not been altered by the introduction of the structural gene.

As used herein, "plant" includes but is not limited to plant cells, plant tissue and plant seeds.

As used herein, "seed endosperm-preferred promoter" is a promoter which preferentially promotes expression of the structural gene in the endosperm of the seed.

As used herein with respect to a structural gene encoding a polypeptide, the term "expresses" means that the structural gene is incorporated into the genome of cells, so that the product encoded by the structural gene is produced within the cells.

As used herein, the term "essential amino acid" means an amino acid which is synthesized only by plants or microorganisms or which is not produced by animals in sufficient quantities to support normal growth and development.

As used herein, the term "high lysine content protein" means that the protein has at least about 7 mole % lysine, preferably about 7 mole % to about 50 mole % lysine, more preferably about 7 mole % to about 40 mole % lysine and most preferably about 7 mole % to about 30 mole %.

As used herein, the term "high sulfur content protein" means that the protein contains at least about 6 mole % methionine and/or cysteine, preferably about 6 mole % to about 40 mole %, more preferably about 6 mole % to about 30 mole % and most preferably 6 mole % to 25 mole %.

5 The present invention provides a transformed plant seed, the endosperm of which is characterized as having an elevated level of a preselected amino acid compared to the seed of a corresponding plant which has not been transformed. It is preferred that the level of preselected amino acid is elevated in the endosperm in preference to other parts of the seed.

10 The preselected amino acid is an essential amino acid such as lysine, cysteine, methionine, threonine, tryptophan, arginine, valine, leucine, isoleucine, histidine or combinations thereof, preferably, the preselected amino acid is lysine, threonine, cysteine, tryptophan, or combinations thereof and optionally methionine. It is especially preferred that the polypeptide has an increased content of lysine as well as a sulfur containing amino acid, i.e., methionine and/or cysteine.

15 The polypeptide can be an endogenous or heterologous protein. When an endogenous protein is expressed, the preselected amino acid is lysine, cysteine, threonine, tryptophan, arginine, valine, leucine, isoleucine, histidine or combinations thereof and optionally methionine. When the protein is a heterologous protein, any of the above described preselected amino acids or combinations thereof is present in elevated amounts.

20 Generally the amount of preselected amino acid in the seed of the present invention is at least about 10 percent by weight greater than in a corresponding untransformed seed, preferably about 10 percent by weight to about 10 times

greater, more preferably about 15 percent by weight to about 10 time greater and most preferably about 20 percent to about 10 times greater.

A polypeptide having an elevated amount of the preselected amino acid is expressed in the transformed plant seed endosperm in an amount sufficient to increase the amount of at least one preselected amino acid in the seed of the transformed plant, relative to the amount of the preselected amino acid in the seed of a corresponding untransformed plant.

The choice of the structural gene is based on the desired amino acid composition of the polypeptide encoded by the structural gene, and the ability of the polypeptide to accumulate in seeds. The amino acid composition of the polypeptide can be manipulated by methods, such as site-directed mutagenesis of the structural gene encoding the polypeptide, so as to result in expression of a polypeptide that is increased in the amount of a particular amino acid. For example, site-directed mutagenesis can be used to increase levels of lysine, methionine, cysteine, threonine and/or tryptophan and/or to decrease levels of asparagine and/or glutamine.

The derivatives differ from the wild-type protein by one or more amino acid substitutions, insertions, deletions or the like. Typically, amino acid substitutions are conservative. In the regions of homology to the native sequence, variants preferably have at least 90% amino acid sequence identity, more preferably at least 95% identity.

Typical examples of suitable proteins include barley chymotrypsin inhibitor, barley alpha hordothionin, soybean 2S albumin proteins, rice high methionine protein and sunflower high methionine protein and derivatives of each protein.

Barley alpha hordothionin has been modified to increase the level of particular amino acids. The sequences of genes which express modified alpha hordothionin proteins with enhanced essential amino acids are based on the mRNA sequence of the native *Hordeum vulgare* alpha hordothionin gene (accession number X05901, Ponz *et al.* 1986 Eur. J. Biochem. 156:131-135).

Modified hordothionin proteins are described in U.S. Ser. Nos. 08/838,763 filed April 10, 1997; 08/824,379 filed March 26, 1997; 08/824,382 filed March 26, 1997; and U.S. Pat. No. 5,703,409 issued December 30, 1997 the disclosures of which are incorporated herein in their entirety by reference.

Alpha hordothionin is a 45-amino acid protein which is stabilized by four disulfide bonds resulting from eight cysteine residues. In its native form, the protein is especially rich in arginine and lysine residues, containing 5 residues (10%) of each. However, it is devoid of the essential amino acid methionine.

Alpha hordothionin has been modified to increase the amount of various amino acids such as lysine, threonine or methionine. The protein has been synthesized and the three-dimensional structure determined by computer modeling. The modeling of the protein predicts that the ten charged residues (arginine at positions 5, 10, 17, 19 and 30, and lysine at positions 1, 23, 32, 38 and 45) all occur on the surface of the molecule. The side chains of the polar amino acids (asparagine at position 11, glutamine at position 22 and threonine at position 41) also occur on the surface of the molecule. Furthermore, the hydrophobic amino acids, (such as the side chains of leucine at positions 8, 15, 24 and 33 and valine at position 18) are also solvent- accessible.

The Three-dimensional modeling of the protein indicates that the arginine residue at position 10 is important to retention of the appropriate 3-dimensional structure and possible folding through hydrogen bond interactions with the C-terminal residue of the protein. A lysine, methionine or threonine substitution at that point would disrupt this hydrogen bonding network, leading to a destabilization of the structure. The synthetic peptide having this substitution could not be made to fold correctly, which supported this analysis. Conservation of the arginine residue at position 10 provides a protein which folds correctly.

Alpha hordothionin has been modified to contain 12 lysine residues in the mature hordothionin peptide, referred to as HT12. (Rao *et al.* 1994 Protein Engineering 7(12):1485-1493 and WO 94/16078 published July 21, 1994) The disclosure of each of these is incorporated herein by reference in their entirety.

Further analysis of substitutions which would not alter the 3-dimensional structure of the molecule led to replacement of Asparagine-11, Glutamine-22 and Threonine-41 with lysine residues with virtually no steric hindrance. The resulting compound contains 27% lysine residues.

Other combinations of these substitutions were also made, including changes in amino acid residues at one or more of positions 5, 11, 17, 19, 22, 30 and 41 are lysine, and the remainder of the residues at those positions are the residues at the corresponding positions in the wild type hordothionin.

Since threonine is a polar amino acid, the surface polar amino acid residues, asparagine at position 11 and glutamine at position 22, can be substituted; and the charged amino acids, lysine at positions 1, 23, 32 and 38 and arginine at positions 5,

17, 19, and 30, can also be substituted with threonine. The molecule can be synthesized by solid phase peptide synthesis.

While the above sequence is illustrative of the present invention, it is not intended to be a limitation. Threonine substitutions can also be performed at positions containing charged amino acids. Only arginine at position 10 and lysine at position 45 are important for maintaining the structure of the protein. One can also substitute at the sites having hydrophobic amino acids. These include positions 8, 15, 18 and 24.

Since methionine is a hydrophobic amino acid, the surface hydrophobic amino acid residues, leucine at positions 8, 15, and 33, and valine at position 18, were substituted with methionine. The surface polar amino acids, asparagine at position 11, glutamine at position 22 and threonine at position 41, are substituted with methionine. The molecule is synthesized by solid phase peptide synthesis and folds into a stable structure. It has seven methionine residues (15.5%) and, including the eight cysteines, the modified protein has a sulfur amino acid content of 33%.

While the above-described proteins are illustrative of suitable polypeptides which can be expressed in the transformed plant, it is not intended to be a limitation. Methionine substitutions can also be performed at positions containing charged amino acids. Only arginine at position 10 is important for maintaining the structure of the protein through a hydrogen-bonding network with serine at position 2 and lysine at position 45. Thus, one can substitute methionine for lysine at positions 1, 23, 32, and/or 38, and for arginine at positions 5, 17, 19 and/or 30.

Many other proteins are also appropriate, for example the protein encoded by the structural gene can be a lysine and/or sulfur rich seed protein, such as the soybean 2S albumin described in U.S. Ser. No. 08/618,911 filed March 20, 1996, and the chymotrypsin inhibitor from barley, Williamson *et al.*, Eur. J Biochem 165: 99-106 (1987), the disclosures of each are incorporated by reference.

Derivatives of these genes can be made by site directed mutagenesis to increase the level of preselected amino acids in the encoded polypeptide. For example the gene encoding for the barley high lysine polypeptide (BHL), is derived from barley chymotrypsin inhibitor, U.S. Ser. No. 08/740,682 filed November 1, 1996 and PCT/US97/20441 filed October 31, 1997, the disclosures of each are incorporated herein by reference. The gene encoding for the enhanced soybean albumin gene (ESA) , is derived from soybean 2S albumin described in U.S. Ser. No. 08/618,911, the disclosure of which is incorporated herein in its entirety by reference.

Other examples of sulfur-rich plant proteins within the scope of the invention include plant proteins enriched in cysteine but not methionine, such as the wheat endosperm purothionine (Mak and Jones; Can. J. Biochem.; Vol. 22; p. 83J; (1976); incorporated herein in its entirety by reference), the pea low molecular weight albumins (Higgins, *et al.*; J. Biol. Chem.; Vol. 261; p. 11124; (1986); incorporated herein in its entirety by reference) as well as 2S albumin genes from other organisms. See, for example, Coulter, *et al.*; J. Exp. Bot.; Vol. 41; p. 1541; (1990); incorporated herein in its entirety by reference.

Such proteins also include methionine-rich plant proteins such as from sunflower seed (Lilley, *et al.*; In: Proceedings of the World Congress on Vegetable Protein Utilization in Human Foods and Animal Feedstuffs; Applewhite, H. (ed.); American Oil Chemists Soc.; Champaign, IL; pp. 497-502; (1989); incorporated
5 herein in its entirety by reference), corn (Pedersen, *et al.*; J. Biol. Chem. p. 261; p. 6279; (1986); Kirihaara, *et al.*; Gene, Vol. 71; p. 359; (1988); both incorporated herein in its entirety by reference), and rice (Musumura, *et al.*; Plant Mol. Biol.; Vol. 12; p. 123; (1989); incorporated herein in its entirety by reference).

The present invention also provides a method for genetically modifying plants
10 to increase the level of at least one preselected amino acid in the endosperm of the seed so as to enhance the nutritional value of the seeds.

The method comprises the introduction of an expression cassette into regenerative plant cells to yield transformed plant cells. The expression cassette comprises a seed endosperm-preferred promoter operably linked to a structural
15 gene encoding a polypeptide elevated in content of the preselected amino acid.

A fertile transformed plant is regenerated from the transformed cells, and seeds are isolated from the plant. The structural gene is transmitted through a complete normal sexual cycle of the transformed plant to the next generation.

The polypeptide is synthesized in the endosperm of seed of the plant which
20 has been transformed by insertion of the expression cassette described above. The sequence for the nucleotide molecule, either RNA or DNA, can readily be derived from the amino acid sequence for the selected polypeptide using standard reference texts.

Plants which can be used in the method of the invention include monocotyledonous cereal plants. Preferred plants include maize, wheat, rice, barley, oats, sorghum, millet and rye. The most preferred plant is maize.

Seeds derived from plants regenerated from transformed plant cells, plant parts or plant tissues, or progeny derived from the regenerated transformed plants, may be used directly as feed or food, or further processing may occur.

Transformation

The transformation of plants in accordance with the invention may be carried out in essentially any of the various ways known to those skilled in the art of plant molecular biology. These include, but are not limited to, microprojectile bombardment, microinjection, electroporation of protoplasts or cells comprising partial cell walls, and *Agrobacterium*-mediated DNA transfer.

I. DNA Used for Transformation

DNA useful for introduction into plant cells includes DNA that has been derived or isolated from any source, that may be subsequently characterized as to structure, size and/or function, chemically altered, and later introduced into the plant.

An example of DNA "derived" from a source, would be a DNA sequence or segment that is identified as a useful fragment within a given organism, and which is then synthesized in essentially pure form. An example of such DNA "isolated" from a source would be a useful DNA sequence that is excised or removed from the source by chemical means, e.g., by the use of restriction endonucleases, so that it can be further manipulated, e.g., amplified, for use in the invention, by the methodology of genetic engineering.

Therefore, useful DNA includes completely synthetic DNA, semi-synthetic DNA, DNA isolated from biological sources, and DNA derived from RNA. The DNA isolated from biological sources, or DNA derived from RNA, includes, but is not limited to, DNA or RNA from plant genes, and non-plant genes such as those from bacteria, yeasts, animals or viruses. The DNA or RNA can include modified genes, portions of genes, or chimeric genes, including genes from the same or different genotype.

The term "chimeric gene" or "chimeric DNA" is defined as a gene or DNA sequence or segment comprising at least two DNA sequences or segments from species which do not recombine DNA under natural conditions, or which DNA sequences or segments are positioned or linked in a manner which does not normally occur in the native genome of untransformed plant.

A structural gene of the invention can be identified by standard methods, e.g., enrichment protocols, or probes, directed to the isolation of particular nucleotide or amino acid sequences. The structural gene can be identified by obtaining and/or screening of a DNA or cDNA library generated from nucleic acid derived from a particular cell type, cell line, primary cells, or tissue.

Screening for DNA fragments that encode all or a portion of the structural gene can be accomplished by screening plaques from a genomic or cDNA library for hybridization to a probe of the structural gene from other organisms or by screening plaques from a cDNA expression library for binding to antibodies that specifically recognize the polypeptide encoded by the structural gene.

DNA fragments that hybridize to a structural gene probe from other organisms and/or plaques carrying DNA fragments that are immunoreactive with antibodies to the polypeptide encoded by the structural gene can be subcloned into a vector and sequenced and/or used as probes to identify other cDNA or genomic sequences encoding all or a portion of the structural gene.

Portions of the genomic copy or copies of the structural gene can be partially sequenced and identified by standard methods including either DNA sequence homology to other homologous genes or by comparison of encoded amino acid sequences to known polypeptide sequences.

Once portions of the structural gene are identified, complete copies of the structural gene can be obtained by standard methods, including cloning or polymerase chain reaction (PCR) synthesis using oligonucleotide primers complementary to the structural gene. The presence of an isolated full-length copy of the structural gene can be verified by comparison of its deduced amino acid sequence with the amino acid sequence of native polypeptide sequences.

As discussed above, the structural gene encoding the polypeptide can be modified to increase the content of particular amino acid residues in that polypeptide by methods well known to the art, including, but not limited to, site-directed mutagenesis. Thus, derivatives of naturally occurring polypeptides can be made by nucleotide substitution of the structural gene so as to result in a polypeptide having a different amino acid at the position in the polypeptide which corresponds to the codon with the nucleotide substitution. The introduction of multiple amino acid changes in a polypeptide can result in a polypeptide which is significantly enriched in a preselected amino acid.

As noted above, the choice of the polypeptide encoded by the structural gene will be based on the amino acid composition of the polypeptide and its ability to accumulate in seeds. The amino acid can be chosen for its nutritional value to produce a value-added trait to the plant or plant part. Amino acids desirable for value-added traits, as well as a source to limit synthesis of an endogenous protein include, but are not limited to, lysine, threonine, tryptophan, methionine, and cysteine.

Expression Cassettes and Expression Vectors

According to the present invention, a structural gene is identified, isolated, and combined with a seed endosperm-preferred promoter to provide a recombinant expression cassette.

The construction of such expression cassettes which can be employed in conjunction with the present invention are well known to those of skill in the art in light of the present disclosure. See, e.g., Sambrook, *et al.*; Molecular Cloning: A Laboratory Manual; Cold Spring Harbor, New York; (1989); Gelvin, *et al.*; Plant Molecular Biology Manual; (1990); Plant Biotechnology: Commercial Prospects and Problems, eds Prakash, *et al.*; Oxford & IBH Publishing Co.; New Delhi, India; (1993); and Heslot, *et al.*; Molecular Biology and Genetic Engineering of Yeasts; CRC Press, Inc., USA; (1992); each incorporated herein in its entirety by reference.

Preferred promoters useful in the practice of the invention are those seed endosperm-preferred promoters that allow expression of the structural gene selectively in seed endosperm to avoid any potential deleterious effects associated with the expression of the structural gene in the embryo.

It has been found that when endosperm-preferred promoters are employed, the total level of the preselected amino acid in the seed is increased compared to a seed produced by employing an embryo-preferred promoter, such as the globulin1 promoter. When the globulin1 promoter is employed, the polypeptide is expressed by the structural gene, but the total amount of the preselected amino acid is not increased.

Examples of suitable promoters include, but are not limited to, 27 kD gamma zein promoter and waxy promoter. See the following sites relating to the 27kD gamma zein promoter: Boronat,A., Martinez,M.C., Reina,M., Puigdomenech,P. and Palau,J.; Isolation and sequencing of a 28 kD glutelin-2 gene from maize: Common elements in the 5' flanking regions among zein and glutelin genes; Plant Sci. 47, 95-102 (1986) and Reina,M., Ponte,I., Guillen,P., Boronat,A. and Palau,J., Sequence analysis of a genomic clone encoding a Zc2 protein from Zea mays W64 A, Nucleic Acids Res. 18 (21), 6426 (1990). See the following site relating to the waxy promoter: Kloesgen,R.B., Gierl,A., Schwarz-Sommer,ZS. and Saedler,H., Molecular analysis of the waxy locus of Zea mays, Mol. Gen. Genet. 203, 237-244 (1986). The disclosures each of these are incorporated herein by reference in their entirety.

However, other endosperm-preferred promoters can be employed.

II. DELIVERY OF DNA TO CELLS

The expression cassette or vector can be introduced into prokaryotic or eukaryotic cells by currently available methods which are described in the literature. See for example, Weising *et al.*, Ann. Rev. Genet. 2: 421-477 (1988). For example, the expression cassette or vector can be introduced into plant cells by methods including, but not limited to, *Agrobacterium*-mediated transformation, electroporation,

PEG poration, microprojectile bombardment, microinjection of plant cell protoplasts or embryogenic callus, silicon fiber delivery, infectious viruses or viroids such as retroviruses, the use of liposomes and the like, all in accordance with well-known procedures.

5 The introduction of DNA constructs using polyethylene glycol precipitation is described in Paszkowski *et al.*, Embo J. 3: 2717-2722 (1984). Electroporation techniques are described in Fromm *et al.*, Proc. Natl. Acad. Sci. 82: 5324 (1985). Ballistic transformation techniques are described in Klein *et al.*, Nature 327: 70-73 (1987). The disclosure of each of these is incorporated herein in its entirety by
10 reference.

 Introduction and expression of foreign genes in plants has been shown to be possible using the T-DNA of the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens*. Using recombinant DNA techniques and bacterial genetics, a wide variety of foreign DNAs can be inserted into T-DNA in *Agrobacterium*. Following
15 infection by the bacterium containing the recombinant Ti plasmid, the foreign DNA is inserted into the host plant chromosomes, thus producing a genetically engineered cell and eventually a genetically engineered plant. A second approach is to introduce root-inducing (Ri) plasmids as the gene vectors.

Agrobacterium tumefaciens-mediated transformation techniques are well
20 described in the literature. See, for example Horsch *et al.*, Science 233: 496-498 (1984), and Fraley *et al.*, Proc. Natl. Acad. Sci. 80: 4803 (1983). *Agrobacterium* transformation of maize is described in U.S. Patent No. 5,550,318. The disclosure of each of these is incorporated herein in its entirety by reference.

Other methods of transfection or transformation include (1) *Agrobacterium rhizogenes*-mediated transformation (see, e.g., Lichtenstein and Fuller In: Genetic Engineering, vol. 6, PWJ Rigby, Ed., London, Academic Press, 1987; and Lichtenstein, C. P., and Draper, J., In: DNA Cloning, Vol. II, D. M. Glover, Ed., Oxford, IRI Press, 1985). Application PCT/US87/02512 (WO 88/02405 published Apr. 7, 1988) describes the use of *A. rhizogenes* strain A4 and its Ri plasmid along with *A. tumefaciens* vectors pARC8 or pARC16 (2) liposome-mediated DNA uptake (see, e.g., Freeman *et al.*, Plant Cell Physiol. 25: 1353, 1984), (3) the vortexing method (see, e.g., Kindle, Proc. Natl. Acad. Sci., USA 87: 1228, (1990). The disclosure of each of these is incorporated herein in its entirety by reference.

DNA can also be introduced into plants by direct DNA transfer into pollen as described by Zhou *et al.*, Methods in Enzymology, 101:433 (1983); D. Hess, Intern Rev. Cytol., 107:367 (1987); Luo *et al.*, Plane Mol. Biol. Reporter, 6:165 (1988). The disclosure of each of these is incorporated herein in its entirety by reference.

Expression of polypeptide coding genes can be obtained by injection of the DNA into reproductive organs of a plant as described by Pena *et al.*, Nature, 325.:274 (1987). The disclosure of which is incorporated herein in its entirety by reference.

DNA can also be injected directly into the cells of immature embryos and the rehydration of desiccated embryos as described by Neuhaus *et al.*, Theor. Appl. Genet., 75:30 (1987); and Benbrook *et al.*, in Proceedings Bio Expo 1986, Butterworth, Stoneham, Mass., pp. 27-54 (1986). The disclosure of each of these is incorporated herein in its entirety by reference.

Plant cells useful for transformation include cells cultured in suspension cultures, callus, embryos, meristem tissue, pollen, and the like.

A variety of plant viruses that can be employed as vectors are known in the art and include cauliflower mosaic virus (CaMV), geminivirus, brome mosaic virus, and tobacco mosaic virus.

Typical vectors useful for expression of genes in higher plants are well known in the art and include vectors derived from the tumor-inducing (Ti) plasmid of *Agrobacterium tumefaciens* described by Rogers *et al.*, Meth. In Enzymol., 153:253-277 (1987). These vectors are plant integrating vectors in that on transformation, the vectors integrate a portion of vector DNA into the genome of the host plant. The disclosure of which is incorporated herein in its entirety by reference.

A particularly preferred vector is a plasmid, by which is meant a circular double-stranded DNA molecule which is not a part of the chromosomes of the cell. Exemplary *A. tumefaciens* vectors useful herein are plasmids pKYLX6 and pKYLX7 of Schardl *et al.*, Gene, 61:1-11 (1987) and Berger *et al.*, Proc. Natl. Acad. Sci. U.S.A., 86:8402-8406 (1989). Another useful vector herein is plasmid pBI101.2 that is available from Clontech Laboratories, Inc. (Palo Alto, CA). The disclosure of each of these is incorporated herein in its entirety by reference.

A cell in which the foreign genetic material in a vector is functionally expressed has been "transformed" by the vector and is referred to as a "transformant".

Either genomic DNA or cDNA coding the gene of interest may be used in this invention. The gene of interest may also be constructed partially from a cDNA clone and partially from a genomic clone.

When the gene of interest has been isolated, genetic constructs are made which contain the necessary regulatory sequences to provide for efficient expression of the gene in the host cell.

According to this invention, the genetic construct will contain (a) a genetic sequence coding for the polypeptide of interest and (b) one or more regulatory sequences operably linked on either side of the structural gene of interest. Typically, the regulatory sequences will be a promoter or a terminator. The regulatory sequences may be from autologous or heterologous sources.

The cloning vector will typically carry a replication origin, as well as specific genes that are capable of providing phenotypic selection markers in transformed host cells. Typically, genes conferring resistance to antibiotics or selected herbicides are used. After the genetic material is introduced into the target cells, successfully transformed cells and/or colonies of cells can be isolated by selection on the basis of these markers.

Typical selectable markers include genes coding for resistance to the antibiotic spectinomycin (e.g., the aada gene), the streptomycin phosphotransferase (SPT) gene coding for streptomycin resistance, the neomycin phosphotransferase (NPTII) gene encoding kanamycin or geneticin resistance, the hygromycin phosphotransferase (HPT) gene coding for hygromycin resistance.

Genes coding for resistance to herbicides include genes which act to inhibit the action of acetolactate synthase (ALS), in particular the sulfonylurea-type herbicides (e.g., the acetolactate synthase (ALS) genes containing mutations leading to such resistance in particular the S4 and/or Hra mutations), genes coding for resistance to herbicides which act to inhibit action of glutamine synthase, such as

phosphinothricin or basta (e.g., the *pat* or *bar* gene), or other such genes known in the art. The *bar* gene encodes resistance to the herbicide basta, and the ALS gene encodes resistance to the herbicide chlorsulfuron.

Typically, an intermediate host cell will be used in the practice of this invention to increase the copy number of the cloning vector. With an increased copy number, the vector containing the gene of interest can be isolated in significant quantities for introduction into the desired plant cells.

Host cells that can be used in the practice of this invention include prokaryotes, including bacterial hosts such as *E. coli*, *S. typhimurium*, and *Serratia marcescens*. Eukaryotic hosts such as yeast or filamentous fungi may also be used in this invention. Since these hosts are also microorganisms, it will be essential to ensure that plant promoters which do not cause expression of the polypeptide in bacteria are used in the vector.

The isolated cloning vector will then be introduced into the plant cell using any convenient transformation technique as described above.

III. Regeneration and Analysis of Transformants

Following transformation, regeneration is involved to obtain a whole plant from transformed cells and the presence of structural gene (s) or "transgene(s)" in the regenerated plant is detected by assays. The seed derived from the plant is then tested for levels of preselected amino acids. Depending on the type of plant and the level of gene expression, introduction of the structural gene into the plant seed endosperm can enhance the level of preselected amino acids in an amount useful to supplement the nutritional quality of those seeds.

Using known techniques, protoplasts and cell or tissue culture can be regenerated to form whole fertile plants which carry and express the gene for a polypeptide according to this invention.

Accordingly, a highly preferred embodiment of the present invention is a transformed maize plant, the cells of which contain at least one copy of the DNA sequence of an expression cassette containing a gene encoding a polypeptide containing elevated amounts of an essential amino acid, such as HT12, BHL or ESA protein.

Techniques for regenerating plants from tissue culture, such as transformed protoplasts or callus cell lines, are known in the art. For example, see Phillips, *et al.*; Plant Cell Tissue Organ Culture; Vol. 1; p. 123; (1981); Patterson, *et al.*; Plant Sci.; Vol. 42; p. 125; (1985); Wright, *et al.*; Plant Cell Reports; Vol. 6; p. 83; (1987); and Barwale, *et al.*; Planta; Vol. 167; p. 473; (1986); each incorporated herein in its entirety by reference. The selection of an appropriate method is within the skill of the art.

Examples of the practice of the present invention detailed herein relate specifically to maize plants. However, the present invention is also applicable to other cereal plants. The expression vectors utilized herein are demonstrably capable of operation in cells of cereal plants both in tissue culture and in whole plants. The invention disclosed herein is thus operable in monocotyledonous species to transform individual plant cells and to achieve full, intact plants which can be regenerated from transformed plant cells and which express preselected polypeptides.

The introduced structural genes are expressed in the transformed plant cells and stably transmitted (somatically and sexually) to the next generation of cells produced. The vector should be capable of introducing, maintaining, and expressing a structural gene in plant cells. The structural gene is passed on to progeny by normal sexual transmission.

To confirm the presence of the structural gene (s) or "transgene(s)" in the regenerating plants, or seeds or progeny derived from the regenerated plant, a variety of assays can be performed. Such assays include Southern and Northern blotting; PCR; assays that detect the presence of a polypeptide product, e.g., by immunological means (ELISAs and Western blots) or by enzymatic function; plant part assays, such as leaf, seed or root assays; and also, by analyzing the phenotype of the whole regenerated plant.

Whereas DNA analysis techniques can be conducted using DNA isolated from any part of a plant, RNA will be expressed in the seed endosperm and hence it will be necessary to prepare RNA for analysis from these tissues.

PCR techniques can be used for detection and quantitation of RNA produced from introduced structural genes. In this application of PCR it is first necessary to reverse transcribe RNA into DNA, using enzymes such as reverse transcriptase, and then through the use of conventional PCR techniques amplify the DNA. In most instances PCR techniques, while useful, will not demonstrate integrity of the RNA product.

Further information about the nature of the RNA product may be obtained by Northern blotting. This technique will demonstrate the presence of an RNA species and give information about the integrity of that RNA. The presence or absence of an

RNA species can also be determined using dot or slot blot Northern hybridizations. These techniques are modifications of Northern blotting and will only demonstrate the presence or absence of an RNA species.

While Southern blotting and PCR may be used to detect the structural gene in question, they do not provide information as to whether the structural gene is being expressed. Expression may be evaluated by specifically identifying the polypeptide products of the introduced structural genes or evaluating the phenotypic changes brought about by their expression.

Assays for the production and identification of specific polypeptides may make use of physical-chemical, structural, functional, or other properties of the polypeptides. Unique physical-chemical or structural properties allow the polypeptides to be separated and identified by electrophoretic procedures, such as native or denaturing gel electrophoresis or isoelectric focusing, or by chromatographic techniques such as ion exchange or gel exclusion chromatography.

The unique structures of individual polypeptides offer opportunities for use of specific antibodies to detect their presence in formats such as an ELISA assay. Combinations of approaches may be employed with even greater specificity such as Western blotting in which antibodies are used to locate individual gene products that have been separated by electrophoretic techniques.

Additional techniques may be employed to absolutely confirm the identity of the product of interest such as evaluation by amino acid sequencing following purification. Although these are among the most commonly employed, other procedures may be additionally used.

Very frequently, the expression of a gene product is determined by evaluating the phenotypic results of its expression. These assays also may take many forms, including but not limited to, analyzing changes in the chemical composition, morphology, or physiological properties of the plant. In particular, the elevated
5 preselected amino acid content due to the expression of structural genes encoding polypeptides can be detected by amino acid analysis.

Breeding techniques useful in the present invention are well known in the art.

The present invention will be further described by reference to the following detailed examples. It is understood, however, that there are many extensions,
10 variations, and modifications on the basic theme of the present invention beyond that shown in the examples and description, which are within the spirit and scope of the present invention.

Examples

EXAMPLE 1

15 ***Construction of the HT12 gene and of other genes encoding polypeptides having an elevated level of a preselected amino acid.***

As noted above, the sequence of the HT12 gene is based on the mRNA sequence of the native *Hordeum vulgare* alpha hordothionin gene (accession number X05901, Ponz *et al.* 1986 Eur. J. Biochem. 156:131-135) modified to
20 introduce 12 lysine residues into the mature hordothionin peptide (See Rao *et al.* 1994 Protein Engineering 7(12):1485-1493, and WO 94/16078 published July 21, 1994).

The alpha hordothionin cDNA comprising the entire alpha hordothionin coding sequence is isolated by rt-PCR of mRNA from developing barley seed. Primers are

designed based upon the published alpha hordothionin sequence to amplify the gene and to introduce a NcoI site at the start (ATG) codon and a BamHI site after the stop codon of the thionin coding sequence to facilitate cloning.

Primers are designated as HTPCR1 (5'-
5 AGTATAAGTAAACACACCATCACACCCTTGAGGCCCTTGCTGGTGGCCATGGT
G-3') and HTPCR2 (5'-

CCTCACATCCCTTAGTGCCTAAGTTCGACGTCGGGCCCTCTAGTCGACGGATC
CA-3'). These primers are used in a PCR reaction to amplify alpha hordothionin by
conventional methods. The resulting PCR product is purified and subcloned into the

10 BamHI/NcoI digested pBSKP vector (Stratagene, LaJolla, CA) and sequenced on
both strands to confirm its identity. The clone is designated pBSKP-HT (seq. ID 1).

Primers are designed for single stranded DNA site-directed mutagenesis to
introduce 12 codons for lysine, based on the peptide structure of hordothionin 12
(Ref: Rao *et al.* 1994 Protein Engineering 7(12):1485-1493) and are designated

15 HT12mut1 (5'-AGCGGAAAATGCCCCGAAAGGCTTCCCCAAATTGGC-3'),
HT12mut2 (5'-

TGCGCAGGCGTCTGCAAGTGTAAGCTGACTAGTAGCGGAAAATGC-3'),

HT12mut3 (5'-

TACAACCTTTGCAAAGTCAAAGGCGCCAAGAAGCTTTGCGCAGGCGTCTG-3'),

20 HT12mut4 (5'-

GCAAGAGTTGCTGCAAGAGTACCCTGGGAAGGAAGTGCTACAACCTTTGC-3').

Sequence analysis is used to verify the desired sequence of the resulting
plasmid, designated pBSKP-HT12 (seq. ID 2).

Similarly, genes encoding other derivatives of hordothionine, as described above, (See U.S. Ser. Nos. 08/838,763 filed April 10, 1997; 08/824,379 filed March 26, 1997; 08/824,382 filed March 26, 1997; and U.S. Pat. No. 5,703,409 issued December 30, 1997), the gene encoding enhanced soybean albumin (ESA) (See U.S. Ser. No. 08/618,911), and genes encoding BHL and other derivatives of the barley chymotrypsin inhibitor (See U.S. Ser. No. 08/740,682 filed November 1, 1996 and PCT/US97/20441 filed October 31, 1997) are constructed by site directed mutagenesis from pBSKP-HT, a subclone of the soybean 2S albumin 3 gene in the pBSKP vector (Stratagene, LaJolla, CA), and a subclone of the barley chymotrypsin inhibitor in the pBSKP vector, respectively.

EXAMPLE 2

Construction of vectors for seed preferred expression of polypeptides having an elevated level of a preselected amino acid.

A 442bp DNA fragment containing the modified hordothionin gene encoding HT12 is isolated from plasmid pBSKP-HT12 by NcoI/BamHI restriction digestion, gel purification and is ligated between the 27 kD gamma zein promoter and 27kD gamma zein terminator of the NcoI/BamHI digested vector PHP3630. PHP 3630 is a subclone of the endosperm-preferred 27kD gamma zein gene (Genbank accession number X58197) in the pBSKP vector (Stratagene), which is modified by site directed mutagenesis by insertion of a NcoI site at the start codon (ATG) of the 27kD gamma zein coding sequence. The 27kD gamma zein coding sequence is replaced with the HT12 coding sequence. The resulting expression vector containing the chimeric gene construct gz::HT12::gz, designated as PHP8001 (Seq. ID 3), is verified by extensive restriction digest analysis and DNA sequencing.

Similarly, the 442bp DNA fragment containing the HT12 coding sequence is inserted between the globulin1 promoter and the globulin1 terminator of the embryo preferred corn globulin1 gene (Genbank accession number X59083), and between the waxy promoter and the waxy terminator of the endosperm-preferred waxy gene (Genbank accession number M24258). The globulin1 and waxy coding sequences, respectively, are replaced with the HT12 coding sequence. The resulting chimeric genes glb1::HT12::glb1, and wx::HT12::wx are designated as PHP 7999 (Seq. ID 4), and PHP 5025 (Seq. ID 5).

In a like manner, expression vectors containing genes encoding other derivatives of hordothionine (See Rao *et al.* 1994 Protein Engineering 7(12):1485-1493, and WO 94/16078 published July 21, 1994), the gene encoding enhanced soybean albumin (ESA) (See U.S. Ser. No. 08/618,911,), and genes encoding BHL and other derivatives of the barley chymotrypsin inhibitor (See U.S. Ser. No. 08/740,682 filed November 1, 1996 and PCT/US97/20441 filed October 31, 1997) are constructed by insertion of the corresponding coding sequences between the promoter and terminator of the 27kD gamma zein gene, the globulin1 gene and the waxy gene, respectively. Resulting chimeric genes are for example gz::ESA::gz and gz::BHL::gz, designated as PHP11260 (Seq. ID 6) and as PHP11427 (Seq. ID 7), respectively.

The resulting expression vectors are used in conjunction with the selectable marker expression cassettes PHP3528 (enhanced CAMV::Bar::PinII) for particle bombardment transformation of maize immature embryos.

EXAMPLE 3

Preparation of Transgenic Plants

The general method of genetic transformation used to produce transgenic maize plants is mediated by bombardment of embryogenically responsive immature embryos with tungsten particles associated with DNA plasmids, said plasmids consisting of a selectable and an unselectable marker gene.

Preparation of Tissue

Immature embryos of "High Type II" are the target for particle bombardment-mediated transformation. This genotype is the F₁ of two purebred genetic lines, parent A and parent B, derived from A188 X B73. Both parents are selected for high competence of somatic embryogenesis. See Armstrong, *et al.*, "Development and Availability of Germplasm with High Type II Culture Formation Response," Maize Genetics Cooperation Newsletter, Vol. 65, pp. 92 (1991); incorporated herein in its entirety by reference.

Ears from F₁ plants are selfed or sibbed, and embryos are aseptically dissected from developing caryopses when the scutellum first becomes opaque. The proper stage occurs about 9-13 days post-pollination, and most generally about 10 days post-pollination, and depends on growth conditions. The embryos are about 0.75 to 1.5 mm long. Ears are surface sterilized with 20-50% Clorox for 30 min, followed by 3 rinses with sterile distilled water.

Immature embryos are cultured, scutellum oriented upward, on embryogenic induction medium comprised of N6 basal salts (Chu, *et al.*, "Establishment of an Efficient Medium for Anther Culture of Rice Through Comparative Experiments on the Nitrogen Sources," Scientia Sinica, (Peking), Vol. 18, pp. 659-668 (1975);

incorporated herein in its entirety by reference; Eriksson vitamins (See Eriksson, T., "Studies on the Growth Requirements and Growth Measurements of Haplopappus gracilis," Physiol. Plant, Vol. 18, pp. 976-993 (1965); incorporated herein in its entirety by reference), 0.5 mg/l thiamine HCl, 30 gm/l sucrose, 2.88 gm/l L-proline, 1 mg/l 2,4-dichlorophenoxyacetic acid, 2 gm/l Gelrite, and 8.5 mg/l AgNO₃.

The medium is sterilized by autoclaving at 121°C for 15 min and dispensed into 100 X 25 mm petri dishes. AgNO₃ is filter-sterilized and added to the medium after autoclaving. The tissues are cultured in complete darkness at 28°C. After about 3 to 7 days, generally about 4 days, the scutellum of the embryo has swelled to about double its original size and the protuberances at the coleorhizal surface of the scutellum indicate the inception of embryogenic tissue. Up to 100% of the embryos display this response, but most commonly, the embryogenic response frequency is about 80%.

When the embryogenic response is observed, the embryos are transferred to a medium comprised of induction medium modified to contain 120 gm/l sucrose. The embryos are oriented with the coleorhizal pole, the embryogenically responsive tissue, upwards from the culture medium. Ten embryos per petri dish are located in the center of a petri dish in an area about 2 cm in diameter. The embryos are maintained on this medium for 3-16 hr, preferably 4 hours, in complete darkness at 28°C just prior to bombardment with particles associated with plasmid DNAs containing the selectable and unselectable marker genes.

To effect particle bombardment of embryos, the particle-DNA agglomerates are accelerated using a DuPont PDS-1000 particle acceleration device. The particle-DNA agglomeration is briefly sonicated and 10 µl are deposited on

macrocarriers and the ethanol allowed to evaporate. The macrocarrier is accelerated onto a stainless-steel stopping screen by the rupture of a polymer diaphragm (rupture disk). Rupture is effected by pressurized helium. Depending on the rupture disk breaking pressure, the velocity of particle-DNA acceleration may be varied. Rupture disk pressures of 200 to 1800 psi are commonly used, with those of 650 to 1100 psi being more preferred, and about 900 psi being most highly preferred. Rupture disk breaking pressures are additive so multiple disks may be used to effect a range of rupture pressures.

Preferably, the shelf containing the plate with embryos is 5.1 cm below the bottom of the macrocarrier platform (shelf #3), but may be located at other distances. To effect particle bombardment of cultured immature embryos, a rupture disk and a macrocarrier with dried particle-DNA agglomerates are installed in the device. The He pressure delivered to the device is adjusted to 200 psi above the rupture disk breaking pressure. A petri dish with the target embryos is placed into the vacuum chamber and located in the projected path of accelerated particles. A vacuum is created in the chamber, preferably about 28 inches Hg. After operation of the device, the vacuum is released and the petri dish is removed.

Bombarded embryos remain on the osmotically adjusted medium during bombardment, and preferably for two days subsequently, although the embryos may remain on this medium for 1 to 4 days. The embryos are transferred to selection medium comprised of N6 basal salts, Eriksson vitamins, 0.5 mg/l thiamine HCl, 30 gm/l sucrose, 1 mg/l 2,4-dichlorophenoxyacetic acid, 2 gm/l Gelrite, 0.85 mg/l AgNO₃ and 3 mg/l bialaphos. Bialaphos is added filter-sterilized. The embryos are subcultured to fresh selection medium at 10 to 14 day intervals. After about 7 weeks,

embryogenic tissue, putatively transgenic for both selectable and unselected marker genes, is seen to proliferate from about 7% of the bombarded embryos. Putative transgenic tissue is rescued, and that tissue derived from individual embryos is considered to be an event and is propagated independently on selection medium.

- 5 Two cycles of clonal propagation is achieved by visual selection for the smallest contiguous fragments of organized embryogenic tissue.

For regeneration of transgenic plants, embryogenic tissue is subcultured to medium comprised of MS salts and vitamins (Murashige, T. and F. Skoog, "A revised medium for rapid growth and bio assays with tobacco tissue cultures";

- 10 Physiologia Plantarum; Vol. 15; pp. 473-497; 1962; incorporated herein in its entirety by reference), 100 mg/l myo-inositol, 60 gm/l sucrose, 3 gm/l Gelrite, 0.5 mg/l zeatin, 1 mg/l indole-3-acetic acid, 26.4 ng/l cis-trans-abscissic acid, and 3 mg/l bialaphos in 100 X 25 mm petri dishes and incubated in darkness at 28°C until the development of well-formed, matured somatic embryos can be visualized. This requires about 14
15 days.

- Well-formed somatic embryos are opaque and cream-colored, and are comprised of an identifiable scutellum and coleoptile. The embryos are individually subcultured to germination medium comprised of MS salts and vitamins, 100 mg/l myo-inositol, 40 gm/l sucrose and 1.5 gm/l Gelrite in 100 X 25 mm petri dishes and
20 incubated under a 16 hr light: 8 hr dark photoperiod and 40 $\mu\text{Einsteinsm}^{-2}\text{sec}^{-1}$ from cool-white fluorescent tubes. After about 7 days, the somatic embryos have germinated and produced a well-defined shoot and root. The individual plants are subcultured to germination medium in 125 x 25 mm glass tubes to allow further plant

development. The plants are maintained under a 16 hr light: 8 hr dark photoperiod and 40 $\mu\text{Einsteinsm}^{-2}\text{sec}^{-1}$ from cool-white fluorescent tubes.

After about 7 days, the plants are well-established and are transplanted to horticultural soil, hardened off, and potted into commercial greenhouse soil mixture and grown to sexual maturity in a greenhouse. An elite inbred line is used as a male to pollinate regenerated transgenic plants.

Preparation of Particles

Fifteen mg of tungsten particles (General Electric) , 0.5 to 1.8 μm , preferably 1 to 1.8 μm , and most preferably 1 μm , are added to 2 ml of concentrated nitric acid.

This suspension is sonicated at 0°C for 20 min (Branson Sonifier Model 450, 40% output, constant duty cycle). Tungsten particles are pelleted by centrifugation at 10,000 rpm (Biofuge) for 1 min and the supernatant is removed. Two ml of sterile distilled water is added to the pellet and sonicate briefly to resuspend the particles. The suspension is pelleted, 1 ml of absolute ethanol is added to the pellet and sonicated briefly to resuspend the particles. Rinse, pellet, and resuspend the particles a further 2 times with sterile distilled water, and finally resuspend the particles in 2 ml of sterile distilled water. The particles are subdivided into 250 μl aliquots and stored frozen.

Preparation of particle-plasmid DNA association

The stock of tungsten particles is sonicated briefly in a water bath sonicator (Branson Sonifier Model 450, 20% output, constant duty cycle) and 50 μl is transferred to a microfuge tube. Plasmid DNA is added to the particles for a final DNA amount of 0.1 to 10 μg in 10 μl total volume, and briefly sonicated. Preferably 1 μg total DNA is used. Specifically, 5 μl of PHP8001 (gz::HT12::gz) and 5 μl of

PHP3528 (enhanced CAMV::Bar::PinII), at 0.1 $\mu\text{g}/\mu\text{l}$ in TE buffer, are added to the particle suspension. Fifty μl of sterile aqueous 2.5 M CaCl_2 are added, and the mixture is briefly sonicated and vortexed. Twenty μl of sterile aqueous 0.1M spermidine are added and the mixture is briefly sonicated and vortexed. The mixture is incubated at room temperature for 20 min with intermittent brief sonication. The particle suspension is centrifuged, and the supernatant is removed. Two hundred fifty μl of absolute ethanol is added to the pellet and briefly sonicated. The suspension is pelleted, the supernatant is removed, and 60 μl of absolute ethanol is added. The suspension is sonicated briefly before loading the particle-DNA agglomeration onto macrocarriers.

EXAMPLE 4

Analysis of seed from transgenic plants for recombinant polypeptides having an elevated level of a preselected amino acid.

Preparation of meals from corn seed

Pooled or individual dry seed harvested from transformed plants from the greenhouse or the field are prepared in one of the following ways:

- A. Seed is imbibed in sterile water overnight (16-20 hr) at 4°C. The imbibed seed is dissected into embryo, endosperm and pericarp. The embryos and endosperm are separately frozen in liquid N_2 , the pericarps are discarded. Frozen tissue is ground with a liquid N_2 chilled ceramic mortar and pestle to a fine meal. The meals are dried under vacuum and stored at -20°C or -80°C.

B. Dry whole seed is ground to a fine meal with a ball mill (Klecko), or by hand with a ceramic mortar and pestle. For analysis of endosperm only, the embryos are removed with a drill and discarded. The remaining endosperm with pericarp is ground with a ball mill or a mortar and pestle.

5 ELISA analysis

Rabbit polyclonal anti HT12 antisera are produced with synthetic HT12 (See Rao *et al. supra*) at Bethyl laboratories. An HT12 ELISA assay is developed and performed by the Analytical Biochemistry department of Pioneer Hi-Bred International, Inc., essentially as described by Harlow and Lane, Antibodies, A Laboratory Manual, Cold Springs Harbor Publication, New York (1988). Quantitative ELISA assays are first performed on pooled meals to identify positive events. Positive events are further analyzed by quantitative ELISA on individual kernels to determine the relative level of HT12 expression and transgene segregation ratio. Among 97 events tested, 59 show HT12 expression levels >1000 ppm. The highest events have HT12 expression levels at 2-5% of the total seed protein. Typical results for HT12 levels for whole kernels of wild type corn, for one event (TC2031) of corn transformed with the gz::HT12::gz chimeric gene, expressing HT12 in the endosperm, for one event (TC320) of corn transformed with the wx::HT12::wx chimeric gene, expressing HT12 in the endosperm, and for one event (TC2027) of corn transformed with the glb1::HT12::glb1 chimeric gene, expressing HT12 in the embryo, are in Table 1.

Similarly, antisera are produced, ELISA assays are developed and assays of seed from transformed plants are performed for other derivatives of hordothionine (See Rao *et al.* 1994 Protein Engineering 7(12):1485-1493, and WO 94/16078 published July 21, 1994), for the enhanced soybean albumin (ESA) (See U.S. Ser.

No. 08/618,911) and for BHL and other derivatives of the barley chymotrypsin inhibitor (See U.S. Ser. No. 08/740,682 filed November 1, 1996 and PCT/US97/20441 filed October 31, 1997), respectively.

Polyacrylamide gel and immuno blot analysis

5 SDS extracts of meals, molecular weight markers, and a synthetic HT12 positive control (see Rao *et al. supra*) are separated on 16.5% or 8-22% polyacrylamide gradient Tris-Tricine gels (Schagger, H. and Von Jagow, G. 1987 Anal. Biochem., 166:368). For immuno blot analysis, gels are transferred to PVDF membranes in 100 mM CAPS, pH 11; 10% methanol using a semidry blotter (Hoefer, San Francisco, CA).

10 After transfer the membrane is blocked in BLOTTO (4% dry milk in Tris-buffered saline, pH 7.5) (Johnson, D. A. , Gausch, J. W., Sportsman, J. R., and Elder, J. H. 1984, Gene Anal. Techn., 1:3). The blots are incubated with rabbit anti-HT12 (same as used for ELISA) diluted 1:2000 to 1:7500 in BLOTTO 2 hr at room temperature (22°C) or overnight at 4°C. Blots are washed 4-5X with BLOTTO, then incubated 1-2

15 hr with horseradish peroxidase-goat anti-rabbit IgG (Promega, Madison, WI) diluted 1:7500 to 1:15000 in BLOTTO. After secondary antibody, the blots are washed 3X with BLOTTO followed by 2 washes with Tris-buffered saline, pH 7.5. Blots are briefly incubated with enhanced chemiluminescence (ECL, Amersham, Arlington Heights, IL) substrate, and wrapped in plastic wrap. Reactive bands are visualized after exposure

20 to x-ray film (Kodak Biomax MR) after short exposure times ranging from 5-120 sec.

HT12 transgenic seed shows a distinctive band not seen in wild type seed at the correct molecular weight and position as judged by the HT12 positive control standard and molecular weight markers. These results indicate that the expressed HT12 prepropeptide is being correctly processed like native HT in barley. Novel polypeptide

bands co-migrating with the HT12 positive control are also observed in Coomassie stained polyacrylamide gels loaded with 10mg total extracted protein indicating substantial expression and accumulation of HT12 protein in the seed.

Similarly, other derivatives of hordothionin, soybean albumin, the enhanced soybean albumin (ESA), BHL and other derivatives of the barley chymotrypsin inhibitor are detected by polyacrylamide gel and immuno blot analysis.

Amino acid composition analysis

Meals from seed, endosperm or embryo that express a recombinant polypeptide having an elevated level of a preselected amino acid are sent to the University of Iowa Protein Structure Facility for amino acid composition analysis using standard protocols for digestion and analysis.

Typical results for the amino acid composition of whole kernels of wild type corn, for one event (TC2031) of corn transformed with the gz::HT12::gz chimeric gene, expressing HT12 in the endosperm, for one event (TC320) of corn transformed with the wx::HT12::wx chimeric gene, expressing HT12 in the endosperm, and for one event (TC2027) of corn transformed with the glb1::HT12::glb1 chimeric gene, expressing HT12 in the embryo, are in Table 1.

Table 1: HT12 ELISA analysis and amino acid composition of meal from whole kernels from wild type corn and from transformed corn expressing recombinant HT12.

transgene none		wx::HT12::wx	gz::HT12::gz	glb1::HT12::glb1
event	wild-type	TC320	TC2031	TC2027
ELISA				
HT 12	protein ppm 0.00	protein ppm 6200	protein ppm 8000	protein ppm 22600
AA				
	Meal % n=3	Meal % n=2	Meal % n=3	Meal % n=4
Lys	0.29	0.38	0.39	0.24
Arg	0.52	0.58	0.56	0.45
Cys	0.12	0.19	0.17	0.22

5 The results in Table 1 demonstrate corn expressing recombinant HT12 in the endosperm shows a significant increase of the preselected amino acid lysine.

Table 2: SEQUENCE INFORMATION

SEQUENCE ID	PROMOTER	GENE
Seq. 1: pBSKP-HT	None	3361-2947
Seq. 2: pBSKP-HT12	None	3361-2947
Seq. 3: PHP8001gz::HT12::gz expression vector	676-2198	2199-2612
Seq. 4: PHP7999 glb1::HT12::glb1 expression vector	3271-1834	1834-1420
Seq. 5: PHP5025 wx::HT::wx expression vector	43-1342	1343-1757
Seq. 6: PHP 11260 gz::ESA::gz expression vector	676-2198	2199-2675
Seq. 7: PHP11427 gz::BHL::gz	676-2198	2199-2450

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Jung, Rudolf
Beach, Larry R.
Dress, Virginia M.
Rao, A. Gururaj
Ranch, Jerome P.
Ertl, David S.
Higgins, Regina K.
- (ii) TITLE OF THE INVENTION: Alteration of Amino Acid Compositions
in Seeds
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Pioneer Hi-Bred International, Inc.
(B) STREET: 7100 NW 62nd Avenue, P.O. Box 1000
(C) CITY: Johnston
(D) STATE: IA
(E) COUNTRY: USA
(F) ZIP: 50131
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Michel, Marianne H
(B) REGISTRATION NUMBER: 35,286
(C) REFERENCE/DOCKET NUMBER: 0815
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 515-334-4467
(B) TELEFAX: 515-334-6883
(C) TELEX:
- (2) INFORMATION FOR SEQ ID NO:1:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3363 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCGACCTCGA	GGGGGGGCCC	GGTACCCAGC	TTTTGTTCCC	TTTAGTGAGG	GTTAATTGCG	60
CGCTTGGCGT	AATCATGGTC	ATAGCTGTTT	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	120
CCACACAACA	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCCTA	ATGAGTGAGC	180
TAACTCACAT	TAATTGCGTT	GCGCTCACTG	CCCCTTTTCC	AGTCGGGAAA	CCTGTCGTGC	240
CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCGTAT	TGGGCGCTCT	300
TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTTCGTT	GGCTGCGGCG	AGCGGTATCA	360
GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	420
ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	480
TTCCATAGGC	TCCGCCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	540
CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	600
TCTCCTGTTC	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCGGGAAGC	660
GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	720
AAGCTGGGCT	GTGTGCACGA	ACCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	780
TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	840
AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	900
AACACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	960
TTTCGAAAAA	GAGTTGGTAG	CTCTTGATCC	GGCAAACAAA	CCACCGCTGG	TAGCGGTGGT	1020
TTTTTTGTTT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	1080
ATCTTTTCTA	CGGGGTCCTGA	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	1140
ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTTTAA	1200
TCAATCTAAA	GTATATATGA	GTAAACTTGG	TCTGACAGTT	ACCAATGCTT	AATCAGTGAG	1260
GCACCTATCT	CAGCGATCTG	TCTATTTCTG	TCATCCATAG	TTGCCTGACT	CCCCGTCGTG	1320
TAGATAACTA	GATACACGGA	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	1380
GACCCACGCT	CACCGGCTCC	AGATTTATCA	GCAATAAACC	AGCCAGCCGG	AAGGGCCGAG	1440
CGCAGAAGTG	GTCCTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	1500
GCTAGAGTAA	GTAGTTCGCC	AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTACAGGC	1560
ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	1620
AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCCCG	1680
ATCGTTGTCA	GAAGTAAGTT	GGCCGCACTG	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	1740
AATTCTCTTA	CTGTCATGCC	ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	GTAATCAACC	1800
AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	CTTGCCCGGC	GTCAATACGG	1860
GATAATACCG	CGCCACATAG	CAGAACTTTA	AAAGTGCTCA	TCATTGGAAA	ACGTTCTTCG	1920
GGGCGAAAAC	TCTCAAGGAT	CTTACCGCTG	TTGAGATCCA	GTTTCGATGA	ACCCACTCGT	1980
GCACCCAACT	GATCTTCAGC	ATCTTTTACT	TTCACGACCG	TTTCTGGGTG	AGCAAAAAACA	2040
GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGCGACAC	GGAAATGTTG	AATACTCATA	2100
CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	2160
ATATTTGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTC	CGCGCACATT	TCCCCGAAAA	2220
GTGCCACCTA	AATTGTAAGC	GTAAATATTT	TGTTAAAAAT	CGCGTTAAAT	TTTTGTATAA	2280
TCAGCTCATT	TTTTAACCAA	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	2340
AGACCGAGAT	AGGGTTGAGT	GTTGTTCCAG	TTTGGAACAA	GAGTCCACTA	TTAAAGAACG	2400
TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	CGATGGCCCA	CTACGTGAAC	2460
CATCACCTTA	ATCAAGTTTT	TTGGGGTCGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCTA	2520
AAGGGAGCCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGGC	GAACGTGGCG	AGAAAGGAAG	2580
GGAAGAAAGC	GAAAGGAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
TAACCACCAC	ACCCGCCGCG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TCGCCATTCA	2700
GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTCGCTATTA	CGCCAGCTGG	2760
CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAA	GCCAGGGTTT	TCCCAGTCAC	2820
GACGTTGTAA	AACGACGGCC	AGTGAGCGCG	CGTAATACGA	CTCACTATAG	GGCGAATTGG	2880
AGCTCCACCG	CGGTGGCGGC	CGCTCTAGAA	CTAGTGGATC	CGTCGACTAG	AGGGCCCCGAC	2940
GTGCAACTTA	GGCACTAAGG	GATGTGAGGC	CAGCATCACC	GTTGCAGAAA	TTGACACAAG	3000
CATCACCACA	ATTTTCCAAA	TAGAGTTTCA	TTTCTTCTGC	GTGAGCAGCT	GCGTTGACCA	3060

TGTAGTCACA	CATGGAAGCC	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGGTTCAT	3120
CTGAGTTGGA	CACAAGGGCC	AATTTGGGGA	AGCCTGTAGG	GCATTTTCCG	CTACTTGTGA	3180
GTTTACACCT	ACAGACGCCT	GCGCATAACT	TCTGAGCACC	ACGGACGCGG	CAAAGGTTGT	3240
AGCAGTTTCT	TCCTAGGGTG	CTCCTGCAGC	AACTCTTGCC	TTCTACTTGC	ACCTGTTCTGA	3300
GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCCTT	GAGGCCCTTG	CTGGTGGCCA	3360
TGG						3363

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3365 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCGACCTCGA	GGGGGGGCCC	GGTACCCAGC	TTTGTGTTCC	TTTAGTGAGG	GTTAATTGCG	60
CGCTTGGCGT	AATCATGGTC	ATAGCTGTTT	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	120
CCACACAACA	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	180
TAACCTACAT	TAATTGCGTT	GCGCTCACTG	CCCGCTTTC	AGTCGGGAAA	CCTGTCGTGC	240
CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCGTAT	TGGGCGCTCT	300
TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTTCGTT	GGCTGCGGCG	AGCGGTATCA	360
GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	420
ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	480
TTCCATAGGC	TCCGCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	540
CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	600
TCTCCTGTTT	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCGGGAAGC	660
GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	720
AAGCTGGGCT	GTGTGCACGA	ACCCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	780
TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	840
AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	900
AACTACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	960
TTCGGAAGAA	GAGTTGGTAG	CTCTTGATCC	GGCAACAAA	CCACCGCTGG	TAGCGGTGGT	1020
TTTTTTTGT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	1080
ATCTTTTCTA	CGGGGTCTGA	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	1140
ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTAAAA	1200
TCAATCTAAA	GTATATATGA	GTAAACTTGG	TCTGACAGTT	ACCAATGCTT	AATCAGTGAG	1260
GCACCTATCT	CAGCGATCTG	TCTATTTCTG	TCTATCCATT	TTGCTGCAAT	CCCCGCTG	1320
TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	1380
GACCCACGCT	CACCGGCTCC	AGATTTATCA	GCAATAAACC	AGCCAGCCGG	AAGGGCCGAG	1440
CGCAGAAGTG	GTCTTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	1500
GCTAGAGTAA	GTAGTTTCGCC	AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTACAGGC	1560
ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	1620
AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCCG	1680
ATCGTTGTCA	GAAGTAAGTT	GGCCGCAAGT	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	1740
AATTCTCTTA	CTGTCAATGC	ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	GTAATCAACC	1800
AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	CTTGCCCGGC	GTCATACCGG	1860
GATAATACCG	CGCCACATAG	CAGAACTTTA	AAAGTGCTCA	TCATTGGAAA	ACGTTCTTCG	1920
GGGCGAAAC	TCTCAAGGAT	CTTACCGCTG	TTGAGATCCA	GTTCGATGTA	ACCCACTCGT	1980
GCACCAACT	GATCTTCAGC	ATCTTTTACT	TTCAACAGCG	TTTCTGGGTG	AGCAAAAAACA	2040
GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	GGAAATGTTG	AATACTCATA	2100
CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	2160
ATATTTGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTT	CGCGCACATT	TCCCCGAAAA	2220
GTGCCACCTA	AATTGTAAGC	GTTAATATTT	TGTTAAAAAT	CGCGTTAAAT	TTTTGTTAAA	2280
TCAGCTCATT	TTTTAACCAA	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	2340

AGACCGAGAT	AGGGTTGAGT	GTTGTTCCAG	TTTGAACAA	GAGTCCACTA	TTAAAGAACG	2400
TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	CGATGGCCCA	CTACGTGAAC	2460
CATCACCCTA	ATCAAGTTTT	TTGGGGTCGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCTA	2520
AAGGGAGCCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGGC	GAACGTGGCG	AGAAAGGAAG	2580
GGAAGAAAGC	GAAAGGAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
TAACCACCAC	ACCCGCCGCG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TCGCCATTCA	2700
GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTCGCTATTA	CGCCAGCTGG	2760
CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAC	GCCAGGGTTT	TCCCAGTCAC	2820
GACGTTGTAA	AACGACGGCC	AGTGAGCGCG	CGTAATACGA	CTCACTATAG	GGCGAATTGG	2880
AGCTCCACCG	CGGTGGCGGC	CGCTCTAGAA	CTAGTGGATC	CGTCGACTAG	AGGGCCCCGAC	2940
GTCGAACTTA	GGCACTAAGG	GATGTGAGGC	CAGCATCACC	GTTGCAGAAA	TTGACACAAG	3000
CATCACCACA	ATTTTCCAAA	TAGAGTTTCA	TTTCTTCGTC	GTCAGCAGCT	GCGTTGACCA	3060
TGTAGTCACA	CATGGAAGCC	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGGTTCAT	3120
CTGAGTTGGA	CACAAGGGCC	AATTTGGGGA	AGCCTTTCCG	GCATTTTCCG	CTACTAGTCA	3180
GCTTACTACT	GCAGACGCC	GCGCAAAGCT	TCTTGGCGCC	TTTGACTTTG	CAAAGGTTGT	3240
AGCACTTCCT	TCCCAGGGTA	CTCTTGCAAG	AACTCTTGCC	TTCTACTTGC	ACCTGTTTCA	3300
GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCCTT	GAGGCCCTTG	CTGGTGGCCA	3360
TGGTG						3365

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5360 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTAAATTGTA	AGCGTTAATA	TTTTGTAA	ATTGCGTTA	AATTTTGT	AAATCAGCTC	60
ATTTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	AAATCAAAAG	AATAGACCGA	120
GATAGGGTTG	AGTGTGTTC	CAGTTTGGA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGTCAAA	GGGCGAAAAA	CCGTCTATCA	GGGCGATGGC	CCACTACGTG	AACCATCACC	240
CTAATCAAGT	TTTTTGGGGT	CGAGGTGCCG	TAAAGCACTA	AATCGGAACC	CTAAAGGGAG	300
CCCCCGATTT	AGAGCTTGAC	GGGGAAAGCC	GGCGAACGTG	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	GCGGGCGCTA	GGGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACCAC	420
CACACCCGCC	GCGCTTAATG	CGCCGCTACA	GGGCGCGTCC	CATTGCGCAT	TCAGGCTGCG	480
CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG	540
GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG	600
TAAACGACG	GCCAGTGAGC	GCGCGTAATA	CGACTCACTA	TAGGGCGAAT	TGGAGCTCCA	660
CCGCGGTGGC	GGCCGCTCTA	GATTATATAA	TTTATAAGCT	AAACAACCCG	GCCCTAAAAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TCGTCGCACG	780
GAATTGCATG	TTTCTTGTG	GAAGCATATT	CACGCAATCT	CCACACATAA	AGGTTTATGT	840
ATAAACTTAC	ATTTAGCTCA	GTTTAATTAC	AGTCTTATTT	GGATGCATAT	GTATGGTTCT	900
CAATCCATAT	AAGTTAGAGT	AAAAAATAAG	TTTAAATTTT	ATCTTAATTC	ACTCCAACAT	960
ATATGGATCT	ACAATACTCA	TGTGCATCCA	AACAAACTAC	TTATATTGAG	GTGAATTTGG	1020
TAGAAATTAA	ACTAACTTAC	ACACTAAGCC	AATCTTTACT	ATATTAAAGC	ACCAGTTTCA	1080
ACGATCGTCC	CGCGTCAATA	TTATTAAAAA	ACTCCTACAT	TTCTTTATAA	TCAACCCGCA	1140
CTCTTATAAT	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGGTGAAA	1200
TTATCTATAA	GTGTTCTGGA	TATTGGTTGT	TGGCTCCCAT	ATTCACACAA	CCTAATCAAT	1260
AGAAAACATA	TGTTTATTAT	AAACAAAATT	TATCATATAT	CATATATATA	TATATATCAT	1320
ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATAACTA	GTGCAACTTA	ATACATGTGT	1380
GTATTAAGAT	GAATAAGAGG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGGATCGAA	1440
AGGGGTTGGA	AACGATTAAA	CGATTAAATC	TCTTCCTAGT	CAAAATTGAA	TAGAAGGAGA	1500
TTTAATATAT	CCCAATCCCC	TTCGATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTGGT	1560
GAGGAACACG	AAAGAACCAT	GCATTGGCAT	GTAAAGCTCC	AAGAATTTGT	TGTATCCTTA	1620

ACAACCTCACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAAG	AAACAATCAA	1680
ACAAATCCTC	TCTGTGTGCA	AAGAAACACG	GTGAGTCACTG	CCGAGATCAT	ACTCATCTGA	1740
TATACATGCT	TACAGCTCAC	AAGACATTAC	AAACAACCTCA	TATTGCATTA	CAAAGATCGT	1800
TTCATGAAAA	ATAAAATAGG	CCGGACAGGA	CAAAAATCCT	TGACGTGTAA	AGTAAATTTA	1860
CAACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCG	TTTTACGTAG	ATCAACAACC	1920
TGTAGAAGGC	AACAAAACCTG	AGCCACGCAG	AAGTACAGAA	TGATTCAGAA	TGAACCATCG	1980
ACGTGCTACG	TAAAGAGAGT	GACGAGTCAT	ATACATTTGG	CAAGAAACCA	TGAAGCTGCC	2040
TACAGCCGTC	TCGGTGCCAT	AAGAACACAA	GAAATTGTGT	TAATTAATCA	AAGCTATAAA	2100
TAACGCTCGC	ATGCCTGTGC	ACTTCTCCAT	CACCACCCT	GGGTCTTCAG	ACCATTAGCT	2160
TTATCTACTC	CAGAGCGCAG	AAGAACCCGA	TCGACACCAT	GGCCACCAGC	AAGGGCCTCA	2220
AGGGTGTGAT	GGTGTGTTTA	CTTATACTGG	GGTTGGTTCT	CGAACAGGTG	CAAGTAGAAG	2280
GCAAGAGTTG	CTGCAAGAGT	ACCCTGGGAA	GGAAGTGCTA	CAACCTTTGC	AAAGTCAAAG	2340
GCGCCAAGAA	GCTTTGCGCA	GGCGTCTGCA	AGTGTAAAGT	GACTAGTAGC	GGAAAATGCC	2400
CGAAAGGCTT	CCCCAAATTG	GCCCTTGTGT	CCAACCTCAG	TGAACCAGAG	ACCGTCAAGT	2460
ATTGCAACTT	GGGGTGTAGG	GCTTCCATGT	GTGACTACAT	GGTCAACGCA	CGTGTGACG	2520
ACGAAGAAAT	GAAACTCTAT	TTGGAAAAAT	GTGGTGATGC	TTGTGTCAAT	TTCTGCAACG	2580
GTGATGCTGG	CCTCACATCC	CTTAGTGCTT	AAGTTCGACG	TCGGGCCCTC	TAGTCGACGG	2640
ATCCCCGGCG	GTGTCCCCCA	CTGAAGAAAC	TATGTGCTGT	AGTATAGCCG	CTGCCCCTG	2700
GCTAGCTAGC	TAGTTGAGTC	ATTTAGCGGC	GATGATTGAG	TAATAATGTG	TCACGCATCA	2760
CCATGCATGG	GTGGCAGTGT	CAGTGTGAGC	AATGACCTGA	ATGAACAATT	GAAATGAAAA	2820
GAAAAAAGTA	TTGTTCCAAA	TTAAACGTTT	TAACCTTTTA	ATAGGTTTAT	ACAATAATTG	2880
ATATATGTTT	TCTGTATATG	TCTAATTTGT	TATCATCCAT	TTAGATATAG	ACAAAAAATA	2940
ATCTAAGAAC	TAAAACAAAT	GCTAATTTGA	AATGAAGGGA	GTATATATTG	GGATAATGTC	3000
GATGAGATCC	CTCGTAATAT	CACCGACATC	ACACGTGTCC	AGTTAATGTA	TCAGTGATAC	3060
GTGTATTAC	ATTTGTTGCG	CGTAGGCGTA	CCCAACAATT	TTGATCGACT	ATCAGAAAGT	3120
CAACGGAAGC	GAGTCGACCT	CGAGGGGGGG	CCCGGTACCC	AGCTTTTGTG	CCCTTTAGTG	3180
AGGGTTAATT	GCGCGCTTGG	CGTAATCATG	GTCATAGCTG	TTTCTGTGT	GAAATTTGTA	3240
TCCGCTCACA	ATTCCACACA	ACATACGAGC	CGGAAGCATA	AAGTGTAAG	CTGGGGTGC	3300
CTAATGAGTG	AGCTAACTCA	CATTAATTGC	GTTGCGCTCA	CTGCCCGCTT	TCCAGTCGGG	3360
AAACCTGTCT	TGCCAGCTGT	ATTAATGAAT	CGGCCAACGC	GCGGGGAGAG	GCGGTTTGCG	3420
TATTGGGCGC	TCTTCCGCTT	CTCTCGCTAC	TGACTCGCTG	CGCTCGGTCT	TTCGGCTGCG	3480
GCGAGCGGTA	TACAGTCACT	CAAAGGCGGT	AATACGGTTA	TCCACAGAAT	CAGGGGATAA	3540
CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	GCAAAAGGCC	AGGAACCGTA	AAAAGGCCGC	3600
GTTGCTGGCG	TTTTTCCATA	GGCTCCGCCC	CCCTGACGAG	CATCACAAAA	ATCGACGCTC	3660
AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	ATAAAGATAC	CAGGCGTTTC	CCCCTGGAAG	3720
CTCCCTCGTG	CGCTCTCCTG	TTCCGACCCT	GCCGCTTACC	GGATACCTGT	CCGCCCTTCT	3780
CCCTTCGGGA	AGCGTGGCGC	TTTCTCATAG	CTCACGCTGT	AGGTATCTCA	GTTCGGTGTA	3840
GGTCGTTTCG	TCCAAGCTGG	GCTGTGTGCA	CGAACCCCC	GTTTCAGCCC	ACCGCTGCGC	3900
CTTATCCGGT	AACTATCGTC	TTGAGTCCAA	CCCGGTAAGA	CACGACTTAT	CGCCACTGGC	3960
AGCAGCCACT	GGTAACAGGA	TTAGCAGAGC	GAGGTATGTA	GGCGGTGCTA	CAGAGTCTCT	4020
GAAGTGTTGG	CCTAACTACG	GCTACACTAG	AAGGACAGTA	TTTGGTATCT	GCGCTCTGCT	4080
GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	TAGCTCTTGA	TCCGGTAAAC	AAACCACCGC	4140
TGGTAGCGGT	GGTTTTTTTT	TTTGCAAGCA	CGCAGATTACG	CGCAGAAAAA	AAGGATCTCA	4200
AGAAGATCCT	TTGATCTTTT	CTACGGGGTC	TGACGCTCAG	TGGAACGAAA	ACTCACGTTA	4260
AGGGATTTTG	GTATGAGAT	TATCAAAAAG	GATCTTCACC	TAGATCCTTT	TAAATTAATA	4320
ATGAAGTTTT	AAATCAATCT	AAAGTATATA	TGAGTAAACT	TGGTCTGACA	GTTACCAATG	4380
CTTAATCAGT	GAGGCACCTA	TCTCAGCGAT	CTGTCTATTT	CGTTTCATCC	TAGTTGCCTG	4440
ACTCCCCGTC	GTGTAGATAA	CTACGATACG	GGAGGGCTTA	CCATCTGGCC	CCAGTGCTGC	4500
AATGATAACG	CGAGACCCAC	GCTCACCAGC	TCCAGATTTA	TCAGCAATAA	ACCAGCCAGC	4560
CGGAAGGGCC	GAGCGCAGAA	GTGGTCCTGC	AACTTTATCC	GCCTCCATCC	AGTCTATTAA	4620
TTGTTGCCCG	GAAGCTAGAG	TAAGTAGTTC	GCCAGTTAAT	AGTTTGCGCA	ACGTTGTTGC	4680
CATTGCTACA	GGCATCGTGG	TGTCACGCTC	GTCGTTTGGT	ATGGCTTCAT	TCAGCTCCGG	4740
TTCCCAACGA	TCAAGGCGAG	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGCTC	4800
CTTCCGTCCT	CCGATCGTTG	TCAGAAGTAA	GTTGGCCGCA	GTGTTATCAC	TCATGTTTAT	4860
GGCAGACTG	CATAATTCTC	TTACTGTGAT	GCCATCCGTA	AGATCGTTTT	CTGTGACTGG	4920
TGAGTACTCA	ACCAAGTCAT	TCTGAGAATA	GTGTATGCGG	CGACCGAGTT	GCTCTTGCCC	4980
GGCGTCAATA	CGGGATAATA	CCGCGCCACA	TAGCAGAACT	TTAAAAGTGC	TCATCATTGG	5040
AAAACGTTCT	TCGGGGCGAA	AACTCTCAAG	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	5100

GTAACCCACT	CGTGCACCCA	ACTGATCTTC	AGCATCTTTT	ACTTTACCA	GCGTTTCTGG	5160
GTGAGCAAAA	ACAGGAAGGC	AAAATGCCGC	AAAAAAGGGA	ATAAGGGCGA	CACGGAAATG	5220
TTGAATACTC	ATACTCTTCC	TTTTTCAATA	TTATTGAAGC	ATTTATCAGG	GTTATTGTCT	5280
CATGAGCGGA	TACATATTG	AATGTATTTA	GAAAAATAAA	CAAATAGGGG	TTCCGCGCAC	5340
ATTTCCCCGA	AAAGTGCCAC					5360

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5511 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TCGCGCGTTT	CGGTGATGAC	GGTGAAAACC	TCTGACACAT	GCAGCTCCCG	GAGACGGTCA	60
CAGCTTGTCT	GTAAGCGGAT	GCCGGGAGCA	GACAAAGCCG	TCAGGGCGCG	TCAGCGGGTG	120
TTGGCGGGTG	TCGGGGCTGG	CTTAACATATG	CGGCATCAGA	GCAGATTGTA	CTGAGAGTGC	180
ACCATATGCG	GTGTGAAATA	CCGCACAGAT	GCGTAAGGAG	AAAATACCGC	ATCAGGCGCC	240
ATTCGCCATT	CAGGCTGCGC	AACTGTTGGG	AAGGGCGATC	GGTGCGGGCC	TCTTCGCTAT	300
TACGCCAGCT	GGCGAAAGGG	GGATGTGCTG	CAAGGCGATT	AAGTTGGGTA	ACGCCAGGGT	360
TTTCCCAGTC	ACGACGTTGT	AAAACGACGG	CCAGTGAATT	CTTTTATGAA	TAATAATAAT	420
GCATATCTGT	GCATTACTAC	CTGGGATACA	AGGGCTTCTC	CGCCATAACA	AATTGAGTTG	480
CGATGCTGAG	AACGAACGGG	GAAGAAAGTA	AGCGCCGCCC	AAAAAAAACG	AACATGTACG	540
TCGGCTATAG	CAGGTGAAAG	TTCGTGCGCC	AATGAAAAGG	GAACGATATG	CGTTGGGTAG	600
TTGGGACTACT	TAAATTTGGA	GAGTTTGTG	CATACACTAA	TCCACTAAAG	TTGTCTATCT	660
TTTTAACAGC	TCTAGGCAGG	ATATAAGATT	TATATCTAAT	CTGTTGGAGT	TGCTTTTAGA	720
GTAACCTTTT	TCTCTGTTTC	GTTTATAGCC	GATTAGCACA	AAATTAAACT	AGGTGACGAG	780
AAATAAAGAA	AAACGGAGGC	AGTAAAAAAT	ACCCAAAAAA	ATACTTGGAG	ATTTTTGTCT	840
CAAAATTATC	TTCTAATTTT	AAAAGCTACA	TATTAATAAT	ACTATATATT	AAAAATACTT	900
CGAGATCATT	GCTTGGGATG	GGCAGGGCCA	ATAGCTAATT	GCTAAGGATG	GGCTATATTT	960
ATGTATCGTC	TGAAACATGT	AGGGGCTAAT	AGTTAGATGA	CTAATTTGCT	GTGTTTCGTAC	1020
GGGGTGCTGT	TTGAGCCTAG	CGATGAAGGG	TCATAGTTTC	ATACAAGAAC	TCACTTTTGG	1080
TTCGTCTGCT	GTGTCTGTTT	TCAGCGTAAC	GGCATCAATG	GATGCCAAAC	TCCGCAAGGG	1140
GACAAATGAA	GAAGCGAAGA	GATTATAGAA	CACGCACGTG	TCATTATTTA	TTTATGGACT	1200
TGCCTCAGTA	GCTTACAGCA	TCGTACCCGC	ACGTACATAC	TACAGAGCCA	CACCTATTGC	1260
ACTGCCTGCC	GCTTACGTAC	ATAGTTAACA	CGCAGAGAGG	TATATACATA	CACGTTCAAC	1320
GTCTCCACTC	AGGCTCATGC	TACGTACGCA	CGTCGGTCGC	GCGCCACCCT	CTCGTTGCTT	1380
CCTGCTCGTT	TTGGCGAGCT	AGAGGGCCCG	ACGTCGAACT	TAGGCACTAA	GGGATGTGAG	1440
GCCAGCATCA	CCGTTGCAGA	AATTGACACA	AGCATCACCA	CAATTTTCCA	AATAGAGTTT	1500
CATTTCTTCG	TCGTACAGCAG	CTGCGTTGAC	CATGTAGTCA	CACATGGAAG	CCCTACACCC	1560
CAAGTTGCAA	TACTTGACGG	TGTCTGGTTC	ATCTGAGTTG	GACACAAGGG	CCAATTTGGG	1620
GAAGCCTTTC	GGGCATTTTC	CGCTACTAGT	CAGCTTACAC	TTGCAGACGC	CTGCGCAAAG	1680
CTTCTTGGCG	CCTTTGACTT	TGCAAAGGTT	GTAGCACTTC	CTTCCCAGGG	TACTCTTGCA	1740
GCAACTCTTG	CCTTCTACTT	GCACCTGTTT	GAGAACCAAC	CCCAGTATAA	GTAACACAC	1800
CATCACACCC	TTGAGGCCCT	TGCTGGTGCC	CATGGTGTA	TGTCGACTGT	GATATCCCTCG	1860
GGTGTGTGTT	GGATCCTTGG	GTTGGCTGTA	TGCAGAACTA	AAGCGGAGGT	GGCGCGCATT	1920
TATACCAGCG	CCGGGCCCTG	GTACGTGGCG	CGGCCGCGCG	GCTACGTGGA	GGAAGGCTGC	1980
GTGGCAGCAG	ACACACGGGT	CGCCACGTCC	CGCCGTACTC	TCCTTACCGT	GCTTATCCCG	2040
GCTCCGGCTC	GGTGACGCCC	AGGGTGTGGC	CGCCTCTGAG	CAGACTTTGT	CGTGTTCCAC	2100
AGTGGTGTCT	TGTTCCGGGG	ACTCCGATCC	GCGGCGAGCG	ACCGAGCGTG	TAAAAGAGTT	2160
CCTACTAGGT	ACGTTTCATTG	TATCTGGACG	ACGGGCAGCG	GACAAATTTG	TGTAAGAGAG	2220
GGGCAGTTTT	TTTTTAGAAA	AACAGAGAAAT	TCCGTTGAGC	TAATTGTAAT	TCAACAAATA	2280
AGCTATTAGT	TGGTTTTAGC	TTAGATTAAA	GAAGCTAACG	ACTAATAGCT	AATAATTAGT	2340
TGGTCTATTA	GTTGACTCAT	TTTAAGGCC	TGTTTCAATC	TCGCGAGATA	AACTTTAGCA	2400

GCTATTTTTT	AGCTACTTTT	AGCCATTTGT	AATCTAAACA	GGAGAGCTAA	TGGTGGTAAT	2460
TGAAACTAAA	CTTTAGCACT	TCAATTCATA	TAGCTAAAGT	TTAGCAGGAA	GCTAAACTTT	2520
ATCCCGTGAG	ATTGAAACGG	GGCCTAAATC	TCTCAGCTAT	TTTTGATGCA	AATTACTGTC	2580
ACTACTGGAA	TCGAGCGCTT	TGCCGAGTGT	CAAAGCCTGA	AAAACACTCC	GTAAAGACTT	2640
TGCCTAGTGT	GACACTCGAC	AAAGAGATCT	CGACGAACAG	TACATCGACA	ACGGCTTCTT	2700
TGTCGAGTAC	TTTTTATCGG	ACACTTGACA	AAGTCTTTGT	CGAGTGAAC	ACATTGAAAC	2760
TCTATGATTT	TATGTGTAGG	TCACTTAGGT	TTCTACACAT	AGTACGTCAC	AACTTTACCG	2820
AAACATTATC	AAATTTTTAT	CACAACCTCT	ATATATGATA	TCATGACATG	TGGACAAGTT	2880
TCATTAATTT	CTGACTTTAT	TTGTGTTTTA	TACAATTTTT	AAACAAC	ATAACAAGTT	2940
CACGGTCATG	TTTAGTGAGC	ATGGTGCTTG	AAGATTTCTG	TCTGCTTCTG	AAATCGGTCTG	3000
TAACCTGTGC	TAGATAACAT	GCATATCATT	TATTTTGCAT	GCACGGTTTT	CCATGTTTCG	3060
AGTGACTTGC	AGTTTAAATG	TGAATTTTCC	GAGAAATTC	AAATAAACGA	ACTAAATCTA	3120
ATATTTATAG	AAAACATTTT	TGTAAATATG	TAATTGTGCC	AAAATGGTAC	ATGTAGATCT	3180
ACATAGTGTA	GGAACATACC	ACAAAAAGTT	TGGTTGGCAA	AATAAAAAAA	ATAAAATATA	3240
CTTTATCGAG	TGTCCAAGGA	TGGCACTCGG	CAAGCTTGGC	GTAATCATGG	TCATAGCTGT	3300
TTCCTGTGTG	AAATTGTTAT	CCGCTCACAA	TTCCACACAA	CATACGAGCC	GGAAGCATAA	3360
AGTGTAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	ATTAATTGCG	TTGCGCTCAC	3420
TGCCCCGCTT	CCAGTCGGGA	AACCTGTCGT	GCCAGCTGCA	TTAATGAATC	GGCCAACGCG	3480
CGGGGAGAGG	CGGTTTTCGT	ATTGGGCGCT	CTTCCGCTTC	CTCGCTCACT	GACTCGCTGC	3540
GCTCGGTTCG	TCGGCTGCGG	CGAGCGGTAT	CAGCTCACTC	AAAGGCGGTA	ATACGGTTAT	3600
CCACAGAATC	AGGGGATAAC	GCAGGAAAGA	ACATGTGAGC	AAAAGGCCAG	CAAAAGGCCA	3660
GGAACCGTAA	AAAGGCCGCG	TTGCTGGCGT	TTTCCCATAG	GCTCCGCCCC	CCTGACGAGC	3720
ATCACAAAAA	TCGACGCTCA	AGTCAGAGGT	GGCGAAACCC	GACAGGACTA	TAAAGATACC	3780
AGGCGTTTCC	CCCTGGAAGC	TCCCTCGTGC	GCTCTCCTGT	TCCGACCCTG	CCGCTTACCG	3840
GATACCTGTC	CGCCTTTCTC	CCTTCGGGAA	GCGTGGCGCT	TTCTCAATGC	TCACGCTGTA	3900
GGTATCTCAG	TTCCGTGTAG	GTCGTTTCGT	CCAAGCTGGG	CTGTGTGCAC	GAACCCCCCG	3960
TTCAGCCCCG	CCGCTGCGCC	TTATCCGGTA	ACTATCGTCT	TGAGTCCAAC	CCGGTAAGAC	4020
ACGACTTATC	GCCACTGGCA	GCAGCCACTG	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	4080
GCGGTGCTAC	AGAGTTCTTG	AAGTGGTGGC	CTAACTACGG	CTACACTAGA	AGGACAGTAT	4140
TTGGTATCTG	CGCTCTGCTG	AAGCCAGTTA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	4200
CCGGCAAACA	AACCACCGCT	GGTAGCGGTG	GTTTTTTTTGT	TTGCAAGCAG	CAGATTACGC	4260
GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TACGGGGTCT	GACGCTCAGT	4320
GGAACGAAAA	CTCACGTTAA	GGGATTTTGG	TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	4380
AGATCCTTTT	AAATTAAAAA	TGAAGTTTTA	AATCAATCTA	AAGTATATAT	GAGTAAACTT	4440
GGTCTGACAG	TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGATC	TGTCTATTTT	4500
GTTTCATCCAT	AGTTGCCTGA	CTCCCCGTCG	TGTAGATAAC	TACGATACGG	GAGGGCTTAC	4560
CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	GAGACCCACG	CTCACC GGCT	CCAGATTTAT	4620
CAGCAATAAA	CCAGCCAGCC	GGAAGGGCCG	AGCGCAGAAG	TGGTCCTGCA	ACTTTATCCG	4680
CCTCCATCCA	GTCTATTAAT	TGTTGCCGGG	AAGCTAGAGT	AAGTAGTTTC	CCAGTTAATA	4740
GTTTGCGCAA	CGTTGTTGCC	ATTGCTACAG	GCATCGTGGT	GTCACGCTCG	TCGTTTGGTA	4800
TGGCTTCATT	CAGCTCCGGT	TCCCAACGAT	CAAGGCGAGT	TACATGATCC	CCCATGTTGT	4860
GCAAAAAAGC	GGTTAGCTCC	TTCCGTCCTC	CGATCGTTGT	CAGAAGTAAG	TTGGCCGCAG	4920
TGTTATCACT	CATGGTTATG	GCAGCACTGC	ATAATTCTCT	TACTGTTCATG	CCATCCGTAA	4980
GATGCTTTTT	TGTGACTGGT	GAGTACTCAA	CCAAGTCATT	CTGAGAATAG	TGTATGCGGC	5040
GACCGAGTTG	CTCTTGCCCG	GCGTCAATAC	GGGATAATAC	CGCGCCACAT	AGCAGAACTT	5100
TAAAAGTGCT	CATCATTGGA	AAACGTTCTT	CGGGGCGAAA	ACTCTCAAGG	ATCTTACCGC	5160
TGTTGAGATC	CAGTTCGATG	TAACCCACTC	GTGCACCCAA	CTGATCTTCA	GCATCTTTTA	5220
CTTTCACCAG	CGTTTCTGGG	TGAGCAAAAA	CAGGAAGGCA	AAATGCCGCA	AAAAAGGGAA	5280
TAAGGGCGAC	ACGGAAATGT	TGAATACTCA	TACTCTTCCT	TTTTCAATAT	TATTGAAGCA	5340
TTTATCAGGG	TTATTGTCTC	ATGAGCGGAT	ACATATTTGA	ATGTATTTAG	AAAAATAAAC	5400
AAATAGGGGT	TCCGCGCACA	TTTCCCCGAA	AAGTGCCACC	TGACGTCTAA	GAAACCATTA	5460
TTATCATGAC	ATTAACCTAT	AAAAATAGGC	GTATCACGAG	GCCCTTTCGT	C	5511

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5115 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTTGGGAGCT	CTCCCATATG	GTCGACCTGC	AGGCGGCCGC	TCTAGAACTA	GTGGATCCCC	60
CCCTCGAGGT	CGACGGTATC	GATAAGCTTG	ATATCTTACA	AGGCCAGCC	CAGCGACCTA	120
TTACACAGCC	CGCTCGGGCC	CGCGACGTCG	GGACACATCT	TCTTCCCCCT	TTTGGTGAAG	180
CTCTGCTCGC	AGCTGTCCGG	CTCCTTGGAG	GTTTCGTGTG	CAGATTCATC	TGTTGTCTCG	240
TCTCCTGTGC	TTCCTGGGTA	GCTTGTGTAG	TGGAGCTGAC	ATGGTCTGAG	CAGGCTTAAA	300
ATTTGCTCGT	AGACGAGGAG	TACCAGCACA	GCACGTTGCG	GATTTCTCTG	CCTGTGAAGT	360
GCAACGTCTA	GGATTGTAC	ACGCCTTGGT	CGCGTCGCGT	CGCGTCGCGT	CGATGCGGTG	420
GTGAGCAGAG	CAGCAACAGC	TGGGCGGCC	AACGTTGGCT	TCCGTGTCTT	CGTCGTACGT	480
ACGCGCGCGC	CGGGGACACG	CAGCAGAGAG	CGGAGAGCGA	GCCGTGCACG	GGGAGGTGGT	540
GTGGAAGTGG	AGCCGCGCGC	CCGGCCGCC	GCGCCCGGTG	GGCAACCCAA	AAGTACCCAC	600
GACAAGCGAA	GGCGCCAAAG	CGATCCAAGC	TCCGGAACGC	AACAGCATGC	GTGCGCTCGG	660
AGAGCCAGCC	ACAAGCAGCC	GAGAACCAG	CCGGTGGGCG	ACGCGTCATG	GGACGGACGC	720
GGGCGACGCT	TCCAAACGGG	CCACGTACGC	CGGCGTGTGC	GTGCGTGCAG	ACGACAAGCC	780
AAGGCGAGGC	AGCCCCGAT	CGGGAAAGCG	TTTTTGGGCG	GAGCGCTGGC	GTGCGGGTCA	840
GTCGCTGGTG	CGCAGTGCCG	GGGGGAACGG	GTATCGTGGG	GGGCGCGGGC	GGAGGAGAGC	900
GTGGCGAGGG	CCGAGAGCAG	CGCGCGGCCG	GGTCACGCAA	CGCGCCCCAC	GTACTGCCCT	960
CCCCCTCCGC	GCGCGCTAGA	AATACCGAGG	CCTGGACCGG	GGGGGGGGCC	CGTCACATCC	1020
ATCCATCGAC	CGATCGATCG	CCACAGCCAA	CACCACCCGC	CGAGGCGACG	CGACAGCCGC	1080
CAGGAGGAAG	GAATAAACTC	ACTGCCAGCC	AGTGAAGGGG	GAGAAGTGTA	CTGCTCCGTC	1140
GACCACTGCG	CGCACCGCCC	GGCAGGGCTG	CTCATCTCGT	CGACGACCAG	GTTCTGTTCC	1200
GATCCGATCC	GATCCTGTCC	TTGAGTTTCG	TCCAGATCCT	GGCGCGTATC	TGCGTGTTTG	1260
ATGATCCAGG	TTCTTCGAAC	CTAAATCTGT	CCGTGCACAC	GTCTTTTCTC	TCTCTCCTAC	1320
GCACTGGATT	AATCGCCATG	GCCACCAGCA	AGGGCCTCAA	GGGTGTGATG	GTGTGTTTAC	1380
TTATACTGGG	GTTGGTTCTC	GAACAGGTGC	AAGTAGAAGG	CAAGAGTTGC	TGCAAGAGTA	1440
CCCTGGGAAG	GAAGTGCTAC	AACCTTTGCA	AAGTCAAAGG	CGCCAAGAAG	CTTTGCGCAG	1500
GCGTCTGCAA	GTGTAAGCTG	ACTAGTAGCG	GAAAATGCCC	GAAAGGCTTC	CCCAAATTGG	1560
CCCTTGTTGTC	CAACTCAGAT	GAACCAGACA	CCGTCAAGTA	TTGCAACTTG	GGGTGTAGGG	1620
CTTCCATGTG	TGACTACATG	GTCAACGCAG	CTGCTGACGA	CGAAGAAATG	AAACTCTATT	1680
TGGAATAATTG	TGGTGATGCT	TGTGTCAATT	TCTGCAACGG	TGATGCTGGC	CTCACATCCC	1740
TTAGTGCCCTA	AGTTTCGACG	CGGGCCCTCT	AGATGCGGCC	CGGGTGAAGA	GTTTCGCCCTG	1800
CAGGGCCCCCT	GATCTCGCGC	GTGGTGCAAA	GATGTTGGGA	CATCTTCTTA	TATATGCTGT	1860
TTTCGCTTATG	TGATATGGAC	AAGTATGTGT	AGATGCTTGC	TTGTGCTAGT	GTAATGTAGT	1920
GTAGTGGTGG	CCAGTGGCAC	AACCTAATAA	GCGCATGAAC	TAATTGCTTG	CGTGTGTAGT	1980
TAAGTACCGA	TCCGTAATTT	TATATTGCGA	GTAAATAAAT	GGACCTGTAG	TGGTGGAGTA	2040
AATAATCCCT	GCTGTTCGGT	GTTCTTATCG	CTCCTCGTAT	AGATATTATA	TAGAGTACAT	2100
TTTTCTCTCT	CTGAATCCTA	CGTGTGTGAA	ATTTCTATAT	CATTACTGTA	AAATTTCTGC	2160
GTTCCAAAAG	AGACCATAGC	CTATCTTTGG	CCCTGTTTTGT	TTTCGGCTTCT	GGCAGCTTCT	2220
GGCCACCAAA	AGCTGCTGCG	GACTGCCAAA	CGCTCAGATT	TTTCAGCTAGC	TTCTATAAAA	2280
TTAGTTGGGG	CAAAAACCAT	CCAAAATCAA	TATAAACACA	TAATCGGTTG	AGTCGTTGTA	2340
ATATTAGGAA	TCTGTCACTT	TCTAGATCCT	GAGCCCTATG	AACAACCTTA	TCTTTCTCCA	2400
TACGTAATCG	TAATGATACT	CAGATTCTCT	CCACAGCCAG	ATTCTCCTCA	CAGCCAGATT	2460
TTCAGAAAAG	CTGGTCAGAA	AAAAGTTAAA	CCAAACAGAC	CCTTTGTGTA	TGCATGGATC	2520
GGCTTTCCCC	GTCAAGCTCT	AAATCGGGGG	CTCCCTTTAG	GGTTCCGATT	TAGAGCTTTA	2580
CGGCACCTCG	ACCGCAAAAA	ACTTGATTTG	GGTGATGGTT	CACGTAGTGG	GCCATCGCCC	2640
TGATAGACGG	TTTTTCGCCC	TTTGACGTTG	GAGTCCACGT	TCTTTAATAG	TGGACTCTTG	2700
TTCCAAACTG	GAACAACACT	CAACCTTATC	TCGGTCTATT	CTTTTGATTT	ATAAGGGATT	2760
TTGCCGATTT	CGGCCTATTG	GTTAAAAAAT	GAGCTGATTT	AACAAATATT	TAACGCGAAT	2820
TTTAACAAAA	TATTAACGTT	TACAAATTCG	CCTGATGCGG	TATTTTCTCC	TTACGCATCT	2880
GTGCGGTATT	TCACACCGCA	TACAGGTGGC	ACTTTTCGGG	GAAATGTGCG	CGGAACCCCT	2940
ATTTGTTTAT	TTTTCTAAAT	ACATTCAAAT	ATGTATCCGC	TCATGAGACA	ATAACCCCTGA	3000
TAAATGCTTC	AATAATATTG	AAAAAGGAAG	AGTATGAGTA	TTCAACATTT	CCGTGTCGCC	3060

CTTATTCCCT	TTTTTGCGGC	ATTTTGCCTT	CCTGTTTTTG	CTCACCCAGA	AACGCTGGTG	3120
AAAGTAAAAG	ATGCTGAAGA	TCAGTTGGGT	GCACGAGTGG	GTTACATCGA	ACTGGATCTC	3180
AACAGCGGTA	AGATCCTTGA	GAGTTTTTCGC	CCCAGAAGAAC	GTTTTTCCAAT	GATGAGCACT	3240
TTTAAAGTTC	TGCTATGTCA	TACACTATTA	TCCCGTATTG	ACGCCGGGCA	AGAGCAAACT	3300
GGTCGCCGGG	CGCGGTATTG	TCAGAATGAC	TTGGTTGAGT	ACTCACCAGT	CACAGAAAAG	3360
CATCTTACGG	ATGGCATGAC	AGTAAGAGAA	TTATGCAGTG	CTGCCATAAC	CATGAGTGAT	3420
AACACTGCGG	CCAACCTACT	TCTGACAACG	ATCGGAGGAC	CGAAGGAGCT	AACCGCTTTT	3480
TTGCACAACA	TGGGGGATCA	TGTAACTCGC	CTTGATCGTT	GGGAACCGGA	GCTGAATGAA	3540
GCCATACCAA	ACGACGAGCG	TGACACCACG	ATGCCTGTAG	CAATGCCAAC	AACGTTGCGC	3600
AAACTATTAA	CTGGCGAACT	ACTTACTCTA	GCTTCCCAGC	AACAATTAAT	AGACTGGATG	3660
GAGGCGGATA	AAGTTGCAGG	ACCACTTCTG	CGCTCGGCCC	TTCCGGCTGG	CTGGTTTATT	3720
GCTGATAAAT	CTGGAGCCGG	TGAGCGTGGG	TCTCGCGGTA	TCATTGCAGC	ACTGGGGCCA	3780
GATGGTAAGC	CCTCCCGTAT	CGTAGTTATC	TACACGACGG	GGAGTCAGGC	AACCTATGGAT	3840
GAACGAAATA	GACAGATCGC	TGAGATAGGT	GCCTCACTGA	TTAAGCATTG	GTAACGTGCA	3900
GACCAAGTTT	ACTCATATAT	ACTTTAGATT	GATTTAAAAAC	TTCAATTTT	ATTTAAAAAGG	3960
ATCTAGGTGA	AGATCCTTTT	TGATAATCTC	ATGACCAAAA	TCCCTTAACG	TGAGTTTTTCG	4020
TTCCACTGAG	CGTCAGACCC	CGTAGAAAAG	ATCAAAGGAT	CTTCTTGAGA	TCCTTTTTTTT	4080
CTGCGCGTAA	TCTGCTGCTT	GCAAACAAAA	AAACCACCGC	TACCAGCGGT	GGTTTTGTTTG	4140
CCGGATCAAG	AGCTACCAAC	TCTTTTTCCG	AAGGTAACCTG	GCTTCAGCAG	AGCGCAGATA	4200
CCAAATACTG	TCCTTCTAGT	GTAGCCGTAG	TTAGGCCACC	ACTTCAAGAA	CTCTGTAGCA	4260
CCGCCTACAT	ACCTCGCTCT	GCTAATCCTG	TTACCAGTGG	CTGCTGCCAG	TGGCGATAAG	4320
TCGTGTCTTA	CCGGGTGGA	CTCAAGACGA	TAGTTACCGG	ATAAGGCGCA	GCGGTCGGGC	4380
TGAACGGGGG	GTTCTGTGCAC	ACAGCCCAGC	TTGGAGCGAA	CGACCTACAC	CGAAGTGAGA	4440
TACCTACAGC	GTGAGCTATG	AGAAAGCGCC	ACGCTTCCCG	AAGGGAGAAA	GGCGGACAGG	4500
TATCCGGTAA	GCGGCAGGGT	CGGAACAGGA	GAGCGCACGA	GGGAGCTTCC	AGGGGGAAAC	4560
GCCTGGTATC	TTTATAGTCC	TGTCGGGTTT	CGCCACCTCT	GACTTGAGCG	TCGATTTTTG	4620
TGATGCTCGT	CAGGGGGGCG	GAGCCTATCG	AAAAACGCCA	GCAACGCGGC	CTTTTTACGG	4680
TTCTTGGCCT	TTTGCTGGCC	TTTTGTCTCAC	ATGTTCTTTC	CTGCGTTATC	CCCTGATTCT	4740
GTGGATAACC	GTATTACCGC	CTTTGAGTGA	GCTGATACCG	CTCGCCGCGC	CCGAACGACC	4800
GAGCGCAGCG	AGTCAGTGAG	CGAGGAAGCG	GAAGAGCGCC	CAATACGCAA	ACCGCCTCTC	4860
CCCGCGCGTT	GGCCGATTCA	TTAATGCAGC	TGGCAGCACA	GGTTTCCCGA	CTGGAAAGCG	4920
GGCAGTGAGC	GCAACGCAAT	TAATGTGAGT	TAGCTCACTC	ATTAGGCACC	CCAGGCTTTA	4980
CACTTTATGC	TTCCGGCTCG	TATGTTGTGT	GGAATTGTGA	GCGGATAACA	ATTTACACACA	5040
GGAAACAGCT	ATGACCATGA	TTACGCCAAG	CTATTTAGGT	GACACTATAG	AATACTCAAG	5100
CTATGCATCC	AACGC					5115

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5392 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTAAATTGTA	AGCGTTAATA	TTTTGTAAAA	ATTCGCGTTA	AATTTTTGTT	AAATCAGCTC	60
ATTTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	AAATCAAAAG	AATAGACCGA	120
GATAGGGTTG	AGTGTTGTTC	CAGTTTGGAA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGTCAAA	GGGCGAAAAA	CCGTCTATCA	GGGCGATGGC	CCACTACGTG	AACCATCACC	240
CTAATCAAGT	TTTTTTGGGGT	CGAGGTGCCG	TAAAGCACTA	AATCGGAACC	CTAAAGGGAG	300
CCCCCGATT	AGAGCTTGAC	GGGGAAAAGC	GGCGAACGTG	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	GCGGGCGCTA	GGGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACCAC	420
CACACCCGCC	GCGCTTAATG	CGCCGCTACA	GGGCGCGTCC	CATTGCGCAT	TCAGGCTGCG	480
CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG	540
GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCACT	CACGACGTTG	600

TAAAACGACG	GCCAGTGAGC	GCGCGTAATA	CGACTCACTA	TAGGGCGAAT	TGGAGCTCCA	660
CCGCGGTGGC	GGCCGCTCTA	GATTATATAA	TTTATAAGCT	AAACAACCCG	GCCCTAAAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TCGTGCGACG	780
GAATTGCATG	TTTCTTGTTG	GAAGCATATT	CACGCAATCT	CCACACATAA	AGGTTTATGT	840
ATAAACTTAC	ATTTAGCTCA	GTTTAATTAC	AGTCTTATTT	GGATGCATAT	GTATGGTTCT	900
CAATCCATAT	AAGTTAGAGT	AAAAAATAAG	TTTAAATTTT	ATCTTAATTC	ACTCCAACAT	960
ATATGGATCT	ACAATACTCA	TGTGCATCCA	AACAACTAC	TTATATTGAG	GTGAATTTGG	1020
TAGAAATTAA	ACTAACTTAC	ACACTAAGCC	AATCTTTACT	ATATTAAAGC	ACCAGTTTCA	1080
ACGATCGTCC	CGCGTCAATA	TTATTAAAAA	ACTCCTACAT	TTCTTTATAA	TCAACCCGCA	1140
CTCTTATAAT	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGGTGAAA	1200
TTATCTATAA	GTGTTCTGGA	TATTGGTTGT	TGGCTCCCAT	ATTCACACAA	CCTAATCAAT	1260
AGAAAACATA	TGTTTTATTA	AAACAAAATT	TATCATATAT	CATATATATA	TATATATCAT	1320
ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATAACTA	GTGCAACTTA	ATACATGTGT	1380
GTATTAAGAT	GAATAAGAGG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGGATCGAA	1440
AGGGGTTGGA	AACGATTAAA	CGATTAAATC	TCTTCCTAGT	CAAAATTGAA	TAGAAGGAGA	1500
TTTAATATAT	CCCAATCCCC	TTCGATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTGGT	1560
GAGGAACACG	AAAGAACCAT	GCATTGGCAT	GTAAAGCTCC	AAGAAATTTGT	TGTATCCTTA	1620
ACAACCTACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAAG	AAACAATCAA	1680
ACAAATCCTC	TCTGTGTGCA	AAGAAACACG	GTGAGTCATG	CCGAGATCAT	ACTCATCTGA	1740
TATACATGCT	TACAGCTCAC	AAGACATTAC	AAACAACCTCA	TATTGCATTA	CAAAGATCGT	1800
TTCATGAAAA	ATAAAATAGG	CCGGACAGGA	CAAAAATCCT	TGACGTGTAA	AGTAAATTTA	1860
CAACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCG	TTTTACGTAG	ATCAACAACC	1920
TGTAGAAGGC	AACAAAACCTG	AGCCACGCAG	AAGTACAGAA	TGATTCCAGA	TGAACCATCG	1980
ACGTGCTACG	TAAAGAGAGT	GACGAGTCAT	ATACATTTGG	CAAGAAACCA	TGAAGCTGCC	2040
TACAGCCGTA	TCGGTGGCAT	AAGAACACAA	GAAATTGTGT	TAATTAATCA	AAGCTATAAA	2100
TAACGCTCGC	ATGCCTGTGC	ACTTCTCCAT	CACCACCAC	GGGTCTTCAG	ACCATTAGCT	2160
TTATCTACTC	CAGAGCGCAG	AAGAACCCGA	TCGACACCAT	GACCAAGTTC	ACAATCCTCC	2220
TCATCTCTCT	TCTCTTCTGC	ATCGCCACCA	CTTGACGCGC	CTCCAAATGG	CAGCACCAGC	2280
AAGATAGCTG	CCGCAAGCAG	CTTAAGGGGG	TGAACCTCAC	GCCCTGCGAG	AAGCACATCA	2340
TGGAGAAGAT	CCAAGGCCGC	GGCGATGACG	ATGATGATGA	TGACGACGAC	AATCACATTC	2400
TCAGGACCAT	GCGGGGGAAG	AATCACTACA	TACGGAAGAA	GGAAGGAAAA	GACGAAGACG	2460
AAGAAGAAGA	AGGACACATG	CAGAAGTGCT	GCGCTTTGCA	CTGGCATTTC	GGGCTCTTAA	2520
GCTCGCTCAT	TTCTGTGCTG	CAGAAGATAA	TGGAGAACCA	GAGCGAGGAA	CTGGAGGAGA	2580
AGGAGAAGAA	GAAAATGGAG	AAGGAGCTTA	TGAACCTGGC	TACTATGTGC	AGGTTTGGGC	2640
CCATGATCGG	GTGCGACTTG	TCCTCCGATG	ACTAAGTTGA	TCCCCGGCGG	TGTCCCCCAC	2700
TGAAGAACT	ATGTGCTGTA	GTATAGCCGC	TGGCTAGCTA	GCTAGTTGAG	TCATTTAGCG	2760
GCGATGATTG	AGTAATAATG	TGTCACGCAT	CACCATGCAT	GGGTGGCAGT	CTCAGTGTGA	2820
GCAATGACCT	GAATGAACAA	TTGAAATGAA	AAGAAAAAAG	TATTGTTCCA	AATTAAACGT	2880
TTTAACCTTT	TAATAGGTTT	ATACAATAAT	TGATATATGT	TTTCTGTATA	TGTCTAATTT	2940
GTTATCATCC	ATTTAGATAT	AGACGAAAAA	AAATCTAAGA	ACTAAAACAA	ATGCTAATTT	3000
GAAATGAAGG	GAGTATATAT	TGGGATAATG	TCGATGAGAT	CCCTCGTAAT	ATCACCGACA	3060
TCACACGTGT	CCAGTTAATG	TATCAGTGAT	ACGTGTATTTC	ACATTTGTTG	CGCGTAGGCG	3120
TACCCAACAA	TTTTGATCGA	CTATCAGAAA	GCTAACGGAA	GCGAGTCGAC	CTCGAGGGGG	3180
GGCCCGGTAC	CCAGCTTTTG	TTCCCTTTAG	TGAGGGTTAA	TTGCGCGCTT	GGCGTAATCA	3240
TGGTCATAGC	TGTTTCCTGT	GTGAAATTGT	TATCCGCTCA	CAATTCCACA	CAACATACGA	3300
GCCGGAAGCA	TAAAGTGTA	AGCCTGGGGT	GCCTAATGAG	TGAGCTAACT	CACATTAATT	3360
GCGTTGCGCT	CACCTGCCGC	TTTCCAGTCG	GGAAACCTGT	CGTGCCAGCT	GCATTAATGA	3420
ATCGGCCAAC	GCGCGGGGAG	AGGCGGTTTG	CGTATTGGGC	GCTCTTCCGC	TTCTTCGCTC	3480
ACTGACTCGC	TGCGCTCGGT	CGTTCCGCTG	CGGCGAGCGG	TATCAGCTCA	CTCAAAGGCG	3540
GTAATACGGT	TATCCACAGA	ATCAGGGGAT	AACGCAGGAA	AGAACATGTG	AGCAAAAGGC	3600
CAGCAAAAGG	CCAGGAACCG	TAAAAAGGCC	GCGTTGCTGG	CGTTTTTCCA	TAGGCTCCGC	3660
CCCCCTGACG	AGCATCACAA	AAATCGACGC	TCAAGTCAGA	GGTGCGGAAA	CCCGACAGGA	3720
CTATAAAGAT	ACCAGGCGTT	TCCCCCTGGA	AGCTCCCTCG	TGCGCTCTCC	TGTTCCGACC	3780
CTGCCGCTTA	CCGGATACCT	GTCCGCTTTT	CTCCCTTCGG	GAAGCGTGGC	GCTTCTCAT	3840
AGCTCACGCT	GTAGGTATCT	CAGTTCGGTG	TAGGTCGTTT	GCTCCAAGCT	GGGCTGTGTG	3900
CACGAACCCC	CCGTTTCAGCC	CGACCGCTGC	GCCTTATCCG	GTAACATATCG	TCTTGAGTCC	3960
AACCCGGTAA	GACACGACTT	ATCGCCACTG	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA	4020
GCGAGGTATG	TAGGCGGTGC	TACAGAGTTC	TTGAAGTGGT	GGCCTAACTA	CGGCTACACT	4080

AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	4140
GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT	TGTTTGCAAG	4200
CAGCAGATTA	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TTCTACGGGG	4260
TCTGACGCTC	AGTGAACGA	AAACTCACGT	TAAGGGATTT	TGGTCATGAG	ATTATCAAAA	4320
AGGATCTTCA	CCTAGATCCT	TTTAAATTAA	AAATGAAGTT	TTAAATCAAT	CTAAAGTATA	4380
TATGAGTAAA	CTTGGTCTGA	CAGTTACCAA	TGCTTAATCA	GTGAGGCACC	TATCTCAGCG	4440
ATCTGTCTAT	TTCGTTTCATC	CATAGTTGCC	TGACTCCCCG	TCGTGTAGAT	AACTACGATA	4500
CGGGAGGGCT	TACCATCTGG	CCCCAGTGCT	GCAATGATAC	CGCGAGACCC	ACGCTCACCG	4560
GCTCCAGATT	TATCAGCAAT	AAACCAGCCA	GCCGGAAGGG	CCGAGCGCAG	AAGTGGTCCT	4620
GCAACTTTAT	CCGCCTCCAT	CCAGTCTATT	AATTGTTGCC	GGGAAGCTAG	AGTAAGTAGT	4680
TCGCCAGTTA	ATAGTTTGCG	CAACGTTGTT	GCCATTGCTA	CAGGCATCGT	GGTGTACACG	4740
TCGTCGTTTT	GTATGGCTTC	ATTGAGTCTC	GGTTCCCAAC	GATCAAGGCG	AGTTACATGA	4800
TCCCCCATGT	TGTGCAAAAA	AGCGGTTAGC	TCCTTCGGTC	CTCCGATCGT	TGTCAGAAGT	4860
AAGTTGGCCG	CAGTGTTATC	ACTCATGGTT	ATGGCAGCAC	TGCATAATTC	TCTTACTGTC	4920
ATGCCATCCG	TAAGATGCTT	TTCTGTGACT	GGTGAGTACT	CAACCAAGTC	ATTCTGAGAA	4980
TAGTGTATGC	GGCGACCGAG	TTGCTCTTGC	CCGGCGTCAA	TACGGGATAA	TACCGCGCCA	5040
CATAGCAGAA	CTTTAAAAGT	GCTCATCATT	GGAAAACGTT	CTTCGGGGCG	AAAACCTCTCA	5100
AGGATCTTAC	CGCTGTTGAG	ATCCAGTTCG	ATGTAACCCA	CTCGTGCACC	CAACTGATCT	5160
TCAGCATCTT	TTACTTTCAC	CAGCGTTTCT	GGGTGAGCAA	AAACAGGAAG	GCAAAATGCC	5220
GCAAAAAAGG	GAATAAGGGC	GACACGGA	TGTTGAATAC	TCATACTCTT	CCTTTTTTCAA	5280
TATTATTGAA	GCATTTATCA	GGGTATTGTT	CTCATGAGCG	GATACATATT	TGAATGTATT	5340
TAGAAAAATA	AACAAATAGG	GGTTCCGCGC	ACATTTCCCC	GAAAAGTGCC	AC	5392

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5173 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTAAATTGTA	AGCGTTAATA	TTTTGTTAAA	ATTCGCGTTA	AATTTTTGTT	AAATCAGCTC	60
ATTTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	AAATCAAAAG	AATAGACCGA	120
GATAGGGTTG	AGTGTGTGTT	CAGTTTGGA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGTCAAA	GGGCGAAAAA	CCGTCTATCA	GGGCGATGGC	CCACTACGTG	AACCATCACC	240
CTAATCAAGT	TTTTTGGGGT	CGAGGTGCCG	TAAAGCACTA	AATCGGAACC	CTAAAGGGAG	300
CCCCGATTT	AGAGCTTGAC	GGGGAAAGCC	GGCGAACGTG	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	GCGGGCGCTA	GGGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACCAC	420
CACACCCGCC	GCGCTTAATG	CGCCGCTACA	GGGCGGTCC	CATTGCGCAT	TCAGGCTGCG	480
CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG	540
GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG	600
TAAACGACG	GCCAGTGAGC	GCGCGTAATA	CGACTCACTA	TAGGGCGAAT	TGGAGCTCCA	660
CCGCGGTGGC	GGCCGCTCTA	GATTATATAA	TTTATAAGCT	AAACAACCCG	GCCCTAAAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TCGTGCGACG	780
GAATTGCATG	TTTCTTGTTG	GAAGCATATT	CACGCAATCT	CCACACATAA	AGGTTTATGT	840
ATAAACTTAC	ATTTAGCTCA	GTTTAATTAC	AGTCTTATTT	GGATGCATAT	GTATGGTTCT	900
CAATCCATAT	AAGTTAGAGT	AAAAAATAAG	TTTAAATTTT	ATCTTAATTC	ACTCCAACAT	960
ATATGGATCT	ACAATACTCA	TGTGCATCCA	AACAAACTAC	TTATATTGAG	GTGAATTTGG	1020
TAGAAATTAA	ACTAACTTAC	ACACTAAGCC	AATCTTTACT	ATATTAAAGC	ACCAGTTTCA	1080
ACGATCGTCC	CGCGTCAATA	TTATTAAAAA	ACTCCTACAT	TTCTTTATAA	TCAACCCGCA	1140
CTCTTATAAT	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGGTGAAA	1200
TTATCTATAA	GTGTTCTGGA	TATTGGTTGT	TGGCTCCCAT	ATTCACACAA	CCTAATCAAT	1260
AGAAAACATA	TGTTTTATTA	AAACAAAATT	TATCATATAT	CATATATATA	TATATATCAT	1320
ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATACTA	GTGCAACTTA	ATACATGTGT	1380

GTATTAAGAT	GAATAAGAGG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGGATCGAA	1440
AGGGGTTGGA	AACGATTAAA	CGATTAAATC	TCTTCCTAGT	CAAAATGTAA	TAGAAGGAGA	1500
TTTAATATAT	CCCAATCCCC	TTGCATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTGGT	1560
GAGGAACACG	AAAGAACCAT	GCATTGGCAT	GTAAGCTCC	AAGAATTTGT	TGTATCCTTA	1620
ACAACTCACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAG	AAACAATCAA	1680
ACAAATCCTC	TCTGTGTGCA	AAGAAACACG	GTGAGTCATG	CCGAGATCAT	ACTCATCTGA	1740
TATACATGCT	TACAGCTCAC	AAGACATTAC	AAACAACCTA	TATTGCATTA	CAAAGATCGT	1800
TTTATGAAAA	ATAAAATAGG	CCGGACAGGA	CAAAAATCCT	TGACGTGTAA	AGTAAATTTA	1860
CAACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCT	TTTTACGTAG	ATCAACAACC	1920
TGTAGAAGGC	AACAAAACTG	AGCCACGCAG	AAGTACAGAA	TGATTCCAGA	TGAACCATCG	1980
ACGTGCTACG	TAAAGAGAGT	GACGAGTCAT	ATACATTTGG	CAAGAAACCA	TGAAGCTGCC	2040
TACAGCCGTA	TCGGTGGCAT	AAGAACACAA	GAAATTGTGT	TAATTAATCA	AAGCTATAAA	2100
TAACGCTCGC	ATGCCTGTGC	ACTTCTCCAT	CACCACCACT	GGGTCTTCAG	ACCATTAGCT	2160
TTATCTACTC	CAGAGCGCAG	AAGAACCCGA	TCGACACCAT	GAAGTCGGTG	GAGAAGAAAAC	2220
CGAAGGGTGT	GAAGACAGGT	GCGGGTGACA	AGCATAAGCT	GAAGACAGAG	TGGCCGGAGT	2280
TGGTGGGGAA	ATCGGTGGAG	AAAGCCAAGA	AGGTGATCCT	GAAGGACAAG	CCAGAGGCGC	2340
AAATCATAGT	TCTACCGGTT	GGTACAAAGG	TGGGTAAGCA	TTATAAGATC	GACAAGGTCA	2400
AGCTTTTTTG	GGATAAAAAG	GACAACATCG	CGCAGGTCCC	CAGGGTCGGC	TAGCCTCGAG	2460
ATCCCCGGCG	GTGTCCCCCA	CTGAAGAAAC	TATGTGCTGT	AGTATAGCCG	CTGGCTAGCT	2520
AGCTAGTTGA	GTCATTTAGC	GGCGATGATT	GAGTAATAAT	GTGTACGCGA	TCACCATGCA	2580
TGGGTGGCAG	TCTCAGTGTG	AGCAATGACC	TGAATGAACA	ATTGAAATGA	AAAGAAAAAA	2640
GTATTGTTCC	AAATTTAAACG	TTTTAACCTT	TTAATGAGGT	TATACAATAA	TTGATATATG	2700
TTTTCTGTAT	ATGTCTAATT	TGTTATCATC	CATTTAGATA	TAGACGAAAA	AAAATCTAAG	2760
AACTAAAACA	AATGCTAATT	TGAAATGAAG	GGAGTATATA	TTGGGATAAT	GTCGATGAGA	2820
TCCCTCGTAA	TATCACCGAC	ATCACACGTG	TCCAGTTAAT	GTATCAGTGA	TACGTGTATT	2880
CACATTTGTT	GCGCGTAGGC	GTACCCAACA	ATTTTGATCG	ACTATCAGAA	AGTCAACGGA	2940
AGCGAGTCGA	CCTCGAGGGG	GGGCCCGGTA	CCCAGCTTTT	GTTCCCTTTA	GTGAGGGTTA	3000
ATTGCGCGCT	TGGCGTAATC	ATGGTCATAG	CTGTTTCCTG	TGTGAAATTG	TTATCCGCTC	3060
ACAATTCAC	ACAACATACG	AGCCGGAAGC	ATAAAGTGTA	AAGCCTGGGG	TGCCTAATGA	3120
GTGAGCTAAC	TCACATTAAT	TGCGTTGCGC	TCACTGCCCG	CTTTCAGTC	GGGAAACCTG	3180
TCGTGCCAGC	TGCATTAATG	AATCGGCCAA	CGCGCGGGGA	GAGGCGGTTT	GCGTATTGGG	3240
CGCTCTTCCG	CTTCTCGCT	CACTGACTCG	CTGCGCTCGG	TCGTTCCGGT	GCGGCGAGCG	3300
GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG	AATCAGGGGA	TAACGCAGGA	3360
AAGAACATGT	GAGCAAAAGG	CCAGCAAAAG	GCCAGGAACC	GTAAAAAGGC	CGCGTTGCTG	3420
GCGTTTTTTC	ATAGGCTCCG	CCCCCTGAC	GAGCATCACA	AAAATCGACG	CTCAAGTCAG	3480
AGGTGGCGAA	ACCCGACAGG	ACTATAAAGA	TACCAAGCGT	TTCCCCCTGG	AAGCTCCCTC	3540
GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT	ACCGGATACC	TGTCCGCCTT	TCTCCCTTCG	3600
GGAAGCGTGG	CGCTTTCTCA	TAGCTCACGC	TGTAGGTATC	TCAGTTCGGT	GTAGGTCGTT	3660
CGCTCCAAGC	TGGGCTGTGT	GCACGAAACC	CCCGTTCAGC	CCGACCGCTG	CGCCTTATCC	3720
GGTAACATATC	GTCTTGAGTC	CAACCCGGTA	AGACACGACT	TATCGCCACT	GGCAGCAGCC	3780
ACTGGTAACA	GGATTAGCAG	AGCGAGGTAT	GTAGGCGGTG	CTACAGAGTT	CTTGAAGTGG	3840
TGGCCTAACT	ACGGCTACAC	TAGAAGGACA	GTATTTGGTA	TCTGCGCTCT	GCTGAAGCCA	3900
GTTACCTTCG	GAAAAAGAGT	TGGTAGCTCT	TGATCCGGCA	AACAAACCAC	CGCTGGTAGC	3960
GGTGGTTTTT	TTGTTTGCAA	GCAGCAGATT	ACGCGCAGAA	AAAAAGGATC	TCAAGAAGAT	4020
CCTTTGATCT	TTTCTACGGG	GTCTGACGCT	CAGTGGAACG	AAAATCACG	TTAAGGGATT	4080
TTGGTCATGA	GATTATCAAA	AAGGATCTTC	ACCTAGATCC	TTTTAAATTA	AAAATGAAGT	4140
TTTAAATCAA	TCTAAAGTAT	ATATGAGTAA	ACTTGGTCTG	ACAGTTACCA	ATGCTTAATC	4200
AGTGAGGCAC	CTATCTCAGC	GATCTGTCTA	TTTCGTTTCAT	CCATAGTTGC	CTGACTCCCC	4260
GTCGTGTAGA	TAACTACGAT	ACGGGAGGGC	TTACCATCTG	GCCCCAGTGC	TGCAATGATA	4320
CCGCGAGACC	CACGCTCACC	GGCTCCAGAT	TTATCAGCAA	TAAACCAGCC	AGCCGGAAGG	4380
GCCGAGCGCA	GAAGTGGTCC	TGCAACTTTA	TCCGCCTCCA	TCCAGCTAT	TAATTGTTGC	4440
CGGGAAGCTA	GAGTAAGTAG	TTCGCCAGTT	AATAGTTTGC	GCAACGTTGT	TGCCATTGCT	4500
ACAGGCATCG	TGGTGTCACG	CTCGTCTGTT	TGATGCGCTT	CATTCAGCTC	CGGTTCCCAA	4560
CGATCAAGGC	GAGTTACATG	ATCCCCCATG	TTGTGCAAAA	AAGCGGTTAG	CTCCTTCGGT	4620
CCTCCGATCG	TTGTTCAGAAG	TAAGTTGGCC	GCAGTGTTAT	CACTCATGGT	TATGGCAGCA	4680
CTGCATAATT	CTCTTACTGT	CATGCCATCC	GTAAGATGCT	TTTCTGTGAC	TGGTGAGTAC	4740
TCAACCAAGT	CATTCTGAGA	ATAGTGTATG				

TCTTCGGGGC	GAAAACTCTC	AAGGATCTTA	CCGCTGTTGA	GATCCAGTTC	GATGTAACCC	4920
ACTCGTGCAC	CCAAC TGATC	TTCAGCATCT	TTTACTTTCA	CCAGCGTTTC	TGGGTGAGCA	4980
AAAACAGGAA	GGCAAAATGC	CGCAAAAAAG	GGAATAAGGG	CGACACGGAA	ATGTTGAATA	5040
CTCATACTCT	TCCTTTTTTC	ATATTATTGA	AGCATTTATC	AGGGTTATTG	TC TCATGAGC	5100
GGATACATAT	TTGAATGTAT	TTAGAAAAAT	AAACAAATAG	GGGTTCCGCG	CACATTTCCC	5160
CGAAAAGTGC	CAC					5173

(2) INFORMATION FOR SEQ ID NO:8:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGCGCAGGCG TCTGCAAGTG TAAGCTGACT AGTAGCGGAA AATGC

45

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAACCTTT GCAAAGTCAA AGGCGCCAAG AAGCTTTGCG CAGGCGTCTG

50

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCAAGAGTTG CTGCAAGAGT ACCCTGGGAA GGAAGTGCTA CAACCTTTGC

50

865020"9T402060

The invention is not limited to the exact details shown and described, for it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention defined by the claims.

09020715.020993

WHAT IS CLAIMED IS

1. A transformed cereal plant seed, the endosperm of which is characterized as having an elevated level of at least one preselected amino acid compared to
5 a seed from a corresponding plant which has not been transformed, wherein the amino acid is lysine, cysteine, threonine, tryptophan, arginine, valine, leucine, isoleucine, histidine or combinations thereof and optionally methionine.
2. The seed according to claim 1 wherein the preselected amino acid is lysine,
10 threonine or tryptophan and optionally a sulfur-containing amino acid.
3. The seed according to Claim 2 wherein the preselected amino acid is lysine.
4. The seed according to Claim 3 wherein the preselected amino acid is lysine and a sulfur-containing amino acid.
5. The seed according to Claim 1 wherein the plant is selected from the group
15 consisting of maize, wheat, rice, barley, oats, sorghum, millet and rye.
6. The seed according to Claim 5 which is a maize seed.
7. The seed according to Claim 1 wherein the plant expresses a transgenic protein having an elevated level of the preselected amino acid.
8. The seed according to Claim 7 wherein the protein is barley chymotrypsin
20 inhibitor, barley alpha hordothionin, soybean 2S albumin protein, rice high methionine protein, sunflower high methionine protein or derivatives of each protein.
9. The seed according to Claim 1 wherein the amount of preselected amino acid in the seed is increased at least about 10 percent by weight compared to a
25 corresponding seed which has not been transformed.

10. The seed according to Claim 9 wherein the amount of the preselected amino acid in the seed is about 10 percent by weight to about 10 times greater compared to a corresponding seed which has not been transformed.

11. The seed according to Claim 10 wherein the amount of the preselected amino acid in the seed is about 15 percent by weight to about 10 times greater compared to a corresponding seed which has not been transformed.

12. The seed according to Claim 11 wherein the amount of the preselected amino acid in the seed is about 20 percent by weight to about 10 times greater compared to a corresponding seed which has not been transformed.

13. An expression cassette comprising a seed endosperm-preferred promoter operably linked to a structural gene encoding a polypeptide elevated in content of a preselected amino acid.

14. The cassette according to Claim 13 wherein the promoter is a gamma zein promoter or a waxy promoter.

15. A vector comprising the expression cassette of Claim 13.

16. A plant cell transformed with the vector of Claim 15.

17. A transformed plant comprising the vector of Claim 15.

18. A seed product obtainable from the transformed seed of Claim 1.

19. A seed from a cereal plant which has been transformed to express a heterologous protein in the endosperm of the seed, wherein the seed exhibits an elevated level of an essential amino acid compared to a plant which has not been transformed.

20. A method for increasing the nutritional value of a cereal plant seed comprising: transforming a host plant cell with a vector comprising an

expression cassette comprising a seed endosperm-preferred promoter operably linked to a structural gene encoding a polypeptide elevated in content of a preselected amino acid; recovering the transformed cells; regenerating a transformed plant; and recovering the seeds therefrom.

- 5 21. A seed produced by the method of claim 20.

09020715 "020999"

ABSTRACT OF THE DISCLOSURE

Rudolf Jung
Larry R. Beach
Virginia M. Dress
A. Gururaj Rao
Jerome P. Ranch
David S. Ertl
Regina K. Higgins

10 The present invention provides a plant seed the endosperm of which is characterized as having an elevated level of a preselected amino acid. The present invention also provides expression cassettes, vectors, plants, plant cells and a method for enhancing the nutritional value of seeds.

CERTIFICATE OF MAILING

I HEREBY CERTIFY THAT THIS CORRESPONDENCE IS BEING DEPOSITED
WITH THE UNITED STATES POSTAL SERVICE AS FIRST-CLASS MAIL IN AN
ENVELOPE ADDRESSED TO: ASSISTANT COMMISSIONER FOR PATENTS,
WASHINGTON, D.C. 20231, ON FEBRUARY 9, 1998

Marianne H. Michel

AGENT/ATTORNEY FOR APPLICANT

February 9, 1998

DATE

Attorney Docket No. 0815

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant: Jung *et al.*

Date: February 9, 1998

Serial No.:

Group Art Unit:

Filed February 9, 1998

Examiner:

For: "ALTERATION OF AMINO ACID COMPOSITIONS IN SEEDS"

Assistant Commissioner for Patents
Box Patent Application
Washington, D.C. 20231

STATEMENT TO SUPPORT FILING AND SUBMISSION IN ACCORDANCE
WITH 37 CFR §§1.821 THROUGH 1.825

- ☒ I hereby state that the contents of the paper and computer readable copies of the Sequence Listing, submitted in accordance with 37 CFR §1.821(c) and (e), respectively, are the same.
- ☐ I hereby state that the submission filed in accordance with 37 CFR §1.821(g) does not include new matter.
- ☐ I hereby state that the submission filed in accordance with 37 CFR §1.821(h) does not include new matter or go beyond the disclosure in the international application as filed.
- ☐ I hereby state that the amendments, made in accordance with 37 CFR §1.825(a), included in the substitute sheet(s) of the Sequence Listing are supported in the application, as filed, at pages _____. I hereby state that the substitute sheet(s) of the Sequence Listing does (do) not include new matter.

09020715-000000

Attorney Docket No.: 0815

- ☐ I hereby state that the substitute copy of the computer readable form, submitted in accordance with 37 CFR §1.825(b), is the same as the amended Sequence Listing.
- ☐ I hereby state that the substitute copy of the computer readable form, submitted in accordance with 37 CFR §1.825(d), is identical to that originally filed.

Respectfully submitted,



Marianne H. Michel
Attorney for Applicant(s)
Registration No. 35,286

PIONEER HI-BRED INTERNATIONAL, INC.
Corporate Intellectual Property
7100 N.W. 62nd Avenue
P.O. Box 1000
Johnston, Iowa 50131-1000
Phone: (515) 334-4467
Facsimile: (515) 334-6883

09020716 020998

INPUT SET: S23914.raw

This Raw Listing contains the General
Information Section and up to the first 5 pages.

SEQUENCE LISTING

ENTERED

1
2
3 (1) General Information
4
5 (i) APPLICANT: Jung, Rudolf
6 Beach, Larry R.
7 Dress, Virginia M.
8 Rao, A. Gururaj
9 Ranch, Jerome P.
10 Ertl, David S.
11 Higgins, Regina K.
12
13 (ii) TITLE OF THE INVENTION: Alteration of Amino Acid Compositions
14 in Seeds
15
16 (iii) NUMBER OF SEQUENCES: 13
17
18 (iv) CORRESPONDENCE ADDRESS:
19 (A) ADDRESSEE: Pioneer Hi-Bred International, Inc.
20 (B) STREET: 7100 NW 62nd Avenue, P.O. Box 1000
21 (C) CITY: Johnston
22 (D) STATE: IA
23 (E) COUNTRY: USA
24 (F) ZIP: 50131
25
26 (v) COMPUTER READABLE FORM:
27 (A) MEDIUM TYPE: Diskette
28 (B) COMPUTER: IBM Compatible
29 (C) OPERATING SYSTEM: DOS
30 (D) SOFTWARE: FastSEQ for Windows Version 2.0
31
32 (vi) CURRENT APPLICATION DATA:
33 (A) APPLICATION NUMBER:
34 (B) FILING DATE:
35 (C) CLASSIFICATION:
36
37 (vii) PRIOR APPLICATION DATA:
38 (A) APPLICATION NUMBER:
39 (B) FILING DATE:
40
41
42
43 (viii) ATTORNEY/AGENT INFORMATION:
44 (A) NAME: Michel, Marianne H
45 (B) REGISTRATION NUMBER: 35,286
46 (C) REFERENCE/DOCKET NUMBER: 0815

09020716.020998

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/020,716DATE: 02/27/98
TIME: 18:53:29

INPUT SET: S23914.raw

47
48 (ix) TELECOMMUNICATION INFORMATION:
49 (A) TELEPHONE: 515-334-4467
50 (B) TELEFAX: 515-334-6883
51 (C) TELEX:
52
53
54 (2) INFORMATION FOR SEQ ID NO:1:
55
56 (i) SEQUENCE CHARACTERISTICS:
57 (A) LENGTH: 3363 base pairs
58 (B) TYPE: nucleic acid
59 (C) STRANDEDNESS: single
60 (D) TOPOLOGY: linear
61
62 (ii) MOLECULE TYPE: Other
63
64 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:
65
66 TCGACCTCGA GGGGGGGGCC GGTACCCAGC TTTTGTTCCT TTTAGTGAGG GTTAATTGCG 60
67 CGCTTGGCGT AATCATGGTC ATAGCTGTTT CCTGTGTGAA ATTGTTATCC GCTCACAATT 120
68 CCACACAACA TACGAGCCGG AAGCATAAAG TGTAAGCCTT GGGGTGCCTA ATGAGTGAGC 180
69 TAACTCACAT TAATTGCGTT GCGCTCACTG CCCGCTTTCC AGTCGGGAAA CCTGTCTGTC 240
70 CAGCTGCATT AATGAATCGG CCAACGCGCG GGGAGAGGCG GTTTGCGTAT TGGGCGCTCT 300
71 TCCGCTTCCT CGCTCACTGA CTCGCTGCGC TCGGTCTGTTT GGCTGCGGCG AGCGGTATCA 360
72 GCTCACTCAA AGGCGGTAAT ACGGTTATCC ACAGAATCAG GGGATAACGC AGGAAAGAAC 420
73 ATGTGAGCAA AAGGCCAGCA AAAGGCCAGG AACCGTAAAA AGGCCGCGTT GCTGGCGTTT 480
74 TTCCATAGGC TCCGCCCCCT TGACGAGCAT CACAAAAATC GACGCTCAAG TCAGAGGTGG 540
75 CGAAACCCGA CAGGACTATA AAGATACCAG GCGTTTCCCT CTGGAAGCTC CCTCGTGCGC 600
76 TCTCCTGTTT CGACCCTGCC GCTTACCGGA TACCTGTCCG CCTTCTCTCC TTCGGGAAGC 660
77 GTGGCGCTTT CTCATAGCTC ACGCTGTAGG TATCTCAGTT CGGTGTAGGT CGTTCGCTCC 720
78 AAGCTGGGCT GTGTGCACGA ACCCCCCGTT CAGCCCGACC GCTGCGCCTT ATCCGGTAAC 780
79 TATCGTCTTG AGTCCAACCC GGTAAGACAC GACTTATCGC CACTGGCAGC AGCCACTGGT 840
80 AACAGGATTA GCAGAGCGAG GTATGTAGGC GGTGCTACAG AGTTCTTGAA GTGGTGGCCT 900
81 AACTACGGCT AACTAGAAG GACAGTATTT GGTATCTGCG CTCTGCTGAA GCCAGTTACC 960
82 TTCGAAAAAA GAGTTGGTAG CTCTTGATCC GGCAACAAA CCACCGCTGG TAGCGGTGGT 1020
83 TTTTTTGTTC GCAAGCAGCA GATTACGCGC AGAAAAAAG GATCTCAAGA AGATCCTTTG 1080
84 ATCTTTTCTA CGGGGTCTGA CGCTCAGTGG AACGAAAAC CACGTTAAGG GATTTTGGTC 1140
85 ATGAGATTAT CAAAAAGGAT CTTACCTAG ATCCTTTTAA ATTAAAAATG AAGTTTAA 1200
86 TCAATCTAAA GTATATATGA GTAAACTTGG TCTGACAGTT ACCAATGCTT AATCAGTGAG 1260
87 GCACCTATCT CAGCGATCTG TCTATTTCTG TCATCCATAG TTGCCTGACT CCCCGTCGTG 1320
88 TAGATAACTA CGATACGGGA GGGCTTACCA TCTGGCCCCA GTGCTGCAAT GATACCGCGA 1380
89 GACCCACGCT CACCGGCTCC AGATTTATCA GCAATAAAC AGCCAGCCGG AAGGGCCGAG 1440
90 CGCAGAAGTG GTCCTGCAAC TTTATCCGCC TCCATCCAGT CTATTAATTG TTGCCGGGAA 1500
91 GCTAGAGTAA GTAGTTTCGCC AGTTAATAGT TTGCGCAACG TTGTTGCCAT TGCTACAGGC 1560
92 ATCGTGGTGT CACGCTCGTC GTTTGGTATG GCTTCATTCA GCTCCGGTTC CCAACGATCA 1620
93 AGGCGAGTTA CATGATCCCC CATGTTGTGC AAAAAAGCGG TTAGCTCCTT CGGTCCTCCG 1680
94 ATCGTTGTCA GAAGTAAGTT GGCCGAGTG TTATCACTCA TGGTTATGGC AGCACTGCAT 1740
95 AATTCTCTTA CTGTCATGCC ATCCGTAAGA TGCTTTTCTG TGAAGTGGTA GTACTCAACC 1800
96 AAGTCATTCT GAGAATAGTG TATGCGGCGA CCGAGTTGCT CTTGCCCGGC GTCAATACGG 1860
97 GATAATACCG CGCCACATAG CAGAACTTTA AAAGTGCTCA TCATTGGAAA ACGTTCTTCG 1920
98 GGGCGAAAAC TCTCAAGGAT CTTACCGCTG TTGAGATCCA GTTCGATGTA ACCCACTCGT 1980
99 GCACCAACT GATCTTCAGC ATCTTTTACT TTCACCAGCG TTTCTGGGTG AGCAAAAACA 2040

09020716.020998

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/020,716

DATE: 02/27/98

TIME: 18:53:32

INPUT SET: S23914.raw

100	GGAAGGCAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	GGAAATGTTG	AATACTCATA	2100
101	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	2160
102	ATATTTGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTT	CGCGCACATT	TCCCCGAAAA	2220
103	GTGCCACCTA	AATTGTAAGC	GTTAATATTT	TGTTAAAAAT	CGCGTTAAAT	TTTTGTAAAA	2280
104	TCAGCTCAT	TTTTAACC	TAAGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAGAAT	2340
105	AGACCGAGAT	AGGGTTGAGT	GTTGTTCCAG	TTTGGAACAA	GAGTCCACTA	TTAAAGAACG	2400
106	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCATCAGGG	CGATGGCCCA	CTACGTGAAC	2460
107	CATCACCCTA	ATCAAGTTT	TTGGGGTCGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCTA	2520
108	AAGGGAGCCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGGC	GAACGTGGCG	AGAAAGGAAG	2580
109	GGAAGAAAGC	GAAAGGAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
110	TAACCACCAC	ACCCGCCGCG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TCGCCATTCA	2700
111	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTCGCTATTA	CGCCAGCTGG	2760
112	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAC	GCCAGGGTTT	TCCCAGTCAC	2820
113	GACGTTGTAA	AACGACGGCC	AGTGAGCGCG	CGTAATACGA	CTCACTATAG	GGCGAATTGG	2880
114	AGCTCCACCG	CGGTGGCGGC	CGCTCTAGAA	CTAGTGGATC	CGTCGACTAG	AGGGCCCCGAC	2940
115	GTCGAACCTA	GGCACTAAGG	GATGTGAGCG	CAGCATCACC	GTTGCAGAAA	TTGACACAAG	3000
116	CATCACACA	ATTTTCCAAA	TAGAGTTTCA	TTTCTTCGTC	GTCAGCAGCT	GCGTTGACCA	3060
117	TGTAGTCACA	CATGGAAGCC	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGGTTTCAT	3120
118	CTGAGTTGGA	CACAAGGGCC	AATTTGGGGA	AGCCTGTAGG	GCATTTTCCG	CTACTTGTGA	3180
119	GTTTACACCT	ACAGACGCC	GCGCATAACT	TCTGAGCACC	ACGGACGCGG	CAAAGGTTGT	3240
120	AGCAGTTTCT	TCCTAGGGTG	CTCCTGCAGC	AACTCTTGCC	TTCTACTTGC	ACCTGTTCTGA	3300
121	GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCCTT	GAGGCCCTTG	CTGGTGGCCA	3360
122	TGG						3363

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3365 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

136	TCGACCTCGA	GGGGGGGGCC	GGTACCCAGC	TTTTGTTC	TTTAGTGAGG	GTAAATTGCG	60
137	CGCTTGGCGT	AATCATGGTC	ATAGCTGTTT	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	120
138	CCACACAACA	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	180
139	TAACCTACAT	TAATTGCGTT	GCGCTCACTG	CCCGCTTTCC	AGTCGGGAAA	CCTGTCGTGC	240
140	CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCCTAT	TGGGCGCTCT	300
141	TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTCGTTT	GGCTGCGGCG	AGCGGTATCA	360
142	GCTCACTCAA	AGGCGGTAAT	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	420
143	ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	480
144	TTCCATAGGC	TCCGCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	540
145	CGAAACCCGA	CAGGACTATA	AAGATACCAG	GCGTTTCCCC	CTGGAAGCTC	CCTCGTGCGC	600
146	TCTCCTGTTT	CGACCCTGCC	GCTTACCGGA	TACCTGTCCG	CCTTTCTCCC	TTCGGGAAGC	660
147	GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTCC	720
148	AAGCTGGGCT	GTGTGCACGA	ACCCCCGTT	CAGCCCGACC	GCTGCGCCTT	ATCCGGTAAC	780
149	TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCGC	CACTGGCAGC	AGCCACTGGT	840
150	AACAGGATTA	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	900
151	AACTACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	960
152	TTCGGAAAAA	GAGTTGGTAG	CTCTTGATCC	GGCAAACAAA	CCACCGCTGG	TAGCGGTGGT	1020

RAW SEQUENCE LISTING
PATENT APPLICATION US/09/020,716

DATE: 02/27/98

TIME: 18:53:35

INPUT SET: S23914.raw

153	TTTTTTGTTT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTG	1080
154	ATCTTTTCTA	CGGGGTCTGA	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	1140
155	ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAATG	AAGTTTAAAA	1200
156	TCAATCTAAA	GTATATATGA	GTAAACTTGG	TCTGACAGTT	ACCAATGCTT	AATCAGTGAG	1260
157	GCACCTATCT	CAGCGATCTG	TCTATTTCTG	TCATCCATAG	TTGCCGTGACT	CCCCGTCTGT	1320
158	TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	GATACCGCGA	1380
159	GACCCACGCT	CACCGGCTCC	AGATTTATCA	GCAATAAACC	AGCCAGCCGG	AAGGGCCGAG	1440
160	CGCAGAAGTG	GTCTTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	1500
161	GCTAGAGTAA	GTAGTTCGCC	AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTACAGGC	1560
162	ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	1620
163	AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCTCC	1680
164	ATCGTTGTCA	GAAGTAAGTT	GGCCGCAGTG	TTATCACTCA	TGGTTATGGC	AGCACTGCAT	1740
165	AATTCCTCTT	CTGTCATGCC	ATCCGTAAGA	TGCTTTTCTG	TGACTGGTGA	GTACTCAACC	1800
166	AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	CTTGCCCGGC	GTCAATACGG	1860
167	GATAATACCG	CGCCACATAG	CAGAACTTTA	AAAGTGCTCA	TCATTGGAAA	ACGTTCTTCG	1920
168	GGGCGAAAAC	TCTCAAGGAT	CTTACCGCTG	TTGAGATCCA	GTTTCGATGT	ACCCACTCGT	1980
169	GCACCCAAC	GATCTTCAGC	ATCTTTTACT	TTCACACGCG	TTTCTGGGTG	AGCAAAAACA	2040
170	GGAAGGCAAA	ATGCCGCAAA	AAAGGGAAAT	AGGGCGACAC	GGAAATGTTG	AATACTCATA	2100
171	CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGTCTCAT	GAGCGGATAC	2160
172	ATATTTGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTT	CGCGCACATT	TCCCCGAAAA	2220
173	GTGCCACCTA	AATTGTAAGC	GTTAATATTT	TGTTAAAATT	CGCGTTAAAT	TTTTTGTTAAA	2280
174	TCAGCTCAT	TTTTAACC	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	2340
175	AGACCGAGAT	AGGGTTGAGT	GTTGTTCCAG	TTTGGAACAA	GAGTCCACTA	TAAAGAACG	2400
176	TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGGG	CGATGGCCCA	CTACGTGAAC	2460
177	CATCACCCTA	ATCAAGTTTT	TTGGGGTCTGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCTA	2520
178	AAGGGAGCCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGGC	GAACGTGGCG	AGAAAGGAAG	2580
179	GGAAGAAAGC	GAAAGGAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
180	TAACCACCAC	ACCCGCCGCG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TCGCCATTCA	2700
181	GGCTGCGCAA	CTGTTGGGAA	GGGCGATCGG	TGCGGGCCTC	TTGCTATTAT	CGCCAGCTGG	2760
182	CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTTGGGTAAC	GCCAGGGTTT	TCCCAGTCAC	2820
183	GACGTTGTAA	AACGACGGCC	AGTGAGCGCG	CGTAATACGA	CTCACTATAG	GGCGAATTGG	2880
184	AGCTCCACCG	CGGTGGCGGG	CGCTCTAGAA	CTAGTGGATC	CGTCGACTAG	AGGGCCCCGAC	2940
185	GTCGAACTTA	GGCACTAAGG	GATGTGAGGC	CAGCATCACC	GTTGCAGAAA	TTGACACAAG	3000
186	CATCACCACA	ATTTTCCAAA	TAGAGTTTCA	TTTCTTCGTC	GTCAGCAGCT	GCGTTGACCA	3060
187	TGTAGTCACA	CATGGAAGCC	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGGTTTCT	3120
188	CTGAGTTGGA	CACAAGGGCC	AATTTGGGGA	AGCCTTTTCGG	GCATTTTCCG	CTACTAGTCA	3180
189	GCTTACACTT	GCAGACGCCT	GCGCAAAGCT	TCTTGGCGCC	TTTGACTTTG	CAAAGGTTGT	3240
190	AGCACTTCCT	TCCCAGGGTA	CTCTTGACGC	AACTCTTGCC	TTCTACTTGC	ACCTGTTCTGA	3300
191	GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCCTT	GAGGCCCTTG	CTGGTGGCCA	3360
192	TGGTG						3365

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5360 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

RAW SEQUENCE LISTING PATENT APPLICATION US/09/020,716

DATE: 02/27/98
TIME: 18:53:39

INPUT SET: S23914.raw

206	CTAAATTGTA	AGCGTTAATA	TTTTGTAA	ATTTCGCGTTA	AATTTTTTGT	AAATCAGCTC	60
207	ATTTTTTAAC	CAATAGGCCG	AAATCGGCAA	AATCCCTTAT	AAATCAAAAG	AATAGACCGA	120
208	GATAGGGTTG	AGTGTGTTC	CAGTTTGGAA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
209	CAACGTCAAA	GGGCGAAAAA	CCGTCTATCA	GGGCGATGGC	CCACTACGTG	AACCATCACC	240
210	CTAATCAAGT	TTTTTGGGGT	CGAGGTGCCG	TAAAGCACTA	AATCGGAACC	CTAAAGGGAG	300
211	CCCCCGATT	AGAGCTTGAC	GGGGAAGCC	GGCGAACGTG	GCGAGAAAGG	AAGGGAAGAA	360
212	AGCGAAAGGA	GCGGGCGCTA	GGGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACCAC	420
213	CACACCCGCC	GCGCTTAATG	CGCCGCTACA	GGGCGCTCC	CATTGCGCAT	TCAGGCTGCG	480
214	CAACTGTTGG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAAGG	540
215	GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCAGT	CACGACGTTG	600
216	TAAAACGACG	GCCAGTGAGC	GCGCGTAATA	CGACTCACTA	TAGGGCGAAT	TGGAGCTCCA	660
217	CCGCGGTGGC	GGCCGCTCTA	GATTATATAA	TTTATAAGCT	AAACAACCCG	GCCCTAAAGC	720
218	ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TCGTGCGCAG	780
219	GAAATGCGATG	TTTCTTGTGT	GAAGCATATT	CACGCAATCT	CCACACATAA	AGGTTTATGT	840
220	ATAAACTTAC	ATTTAGCTCA	GTTTAATTAC	AGTCTTATTT	GGATGCATAT	GTATGGTTCT	900
221	CAATCCATAT	AAGTTAGAGT	AAAAAATAAG	TTTAAATTTT	ATCTTAATTC	ACTCCAACAT	960
222	ATATGGATCT	ACAATACTCA	TGTGCATCCA	AACAAACTAC	TTATATTGAG	GTGAATTTGG	1020
223	TAGAAATFAA	ACTAACTTAC	ACACTAAGCC	AATCTTTACT	ATATTAAAGC	ACCAGTTTCA	1080
224	ACGATCGTCC	CGCGTCAATA	TTATTAAAAA	ACTCCTACAT	TTCTTTATAA	TCAACCCGCA	1140
225	CTCTTATAAT	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGGTGAAA	1200
226	TTATCTATAA	GTGTTCTGGA	TATTTGGTTGT	TGGCTCCCAT	ATTCACACAA	CCTAATCAAT	1260
227	AGAAAACATA	TGTTTTATTA	AAACAAAATT	TATCATATAT	CATATATATA	TATATATCAT	1320
228	ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATAACTA	GTGCAACTTA	ATACATGTGT	1380
229	GTATTAAAGAT	GAATAAGAGG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGGATCGAA	1440
230	AGGGGTTGGA	AACGATTAAA	CGATTAAATC	TC'TTCC'TAGT	CAAAATTGAA	TAGAAGGAGA	1500
231	TTTAATATAT	CCCAATCCCC	TTCGATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTTGGT	1560
232	GAGGAACACG	AAAGAACCAT	GCATTGGCAT	GTAAAGCTCC	AAGAATTTGT	TGTATCCTTA	1620
233	ACAACTCACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAG	AAACAATCAA	1680
234	ACAAATCCTC	TCTGTGTGCA	AAGAAACACG	GTGAGTCATG	CCGAGATCAT	ACTCATCTGA	1740
235	TATACATGCT	TACAGCTCAC	AAGACATTAC	AAACAAC'TCA	TAT'TGCATTA	CAAAGATCGT	1800
236	TTCATGAAAA	ATAAAATAGG	CCGGACAGGA	CAAAAATCCT	TGACGTGTAA	AGTAAATTTA	1860
237	CAACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCG	TTTTACGTAG	ATCAACAACC	1920
238	TGTAGAAGGC	AACAAAAC'TG	AGCCACGCAG	AAGTACAGAA	TGATTCCAGA	TGAACCATCG	1980
239	ACGTGCTACG	TAAAGAGAGT	GACGAGTCAT	ATACATTTGG	CAAGAAACCA	TGAAGCTGCC	2040
240	TACAGCCGTC	TCGGTGGCAT	AAGAACACAA	GAAAT'TGTGT	TAATTAATCA	AAGCTATAAA	2100
241	TAACGCTCGC	ATGCC'TGTGC	ACTTCTCCAT	CACCACCACT	GGGTCTTCAG	ACCATTAGCT	2160
242	TTATCTACTC	CAGAGCGCAG	AAGAACCCGA	TGCACACCAT	GGCCACCAGC	AAGGGCCTCA	2220
243	AGGGTGTGAT	GGTGTGTTTA	CTTATAC'TGG	GGTTGGTTCT	CGAACAGGTG	CAAGTAGAAG	2280
244	GCAAGAGTTG	CTGCAAGAGT	ACCC'TGGGAA	GGAAGTGTCTA	CAACCTTTGC	AAAGTCAAAG	2340
245	GCGCCAAGAA	GCTTTGCGCA	GGCGTCTGCA	AGTGTAAAGCT	GACTAGTAGC	GGAAAATGCC	2400
246	CGAAAGGCTT	CCCCAAATTG	GCCCTTGTGT	CCAAC'TCAGA	TGAACCAGAC	ACCGTCAAGT	2460
247	ATTGCAACTT	GGGGTGTAGG	GCTTCCATGT	GTGACTACAT	GGTCAACGCA	GCTGCTGACG	2520
248	ACGAAGAAAT	GAAAC'TCTAT	TTGGAAAATT	GTGGTGTATGC	TTGTGTCAAT	TTCTGCAACG	2580
249	GTGATGCTGG	CTC'CATCC	CTTAGTGCCT	AAGTTCGACG	TCGGGCCCTC	TAGTCGACGG	2640
250	ATCCCCGGCG	GTGTCCCCCA	CTGAAGAAAC	TATGTGCTGT	AGTATAGCCG	CTGCCCGCTG	2700
251	GCTAGCTAGC	TAGTTGAGTC	ATTTAGCGGC	GATGATTGAG	TAATAATGTG	TCACGCATCA	2760
252	CCATGCATGG	GTGGCAGTGT	CAGTGTGAGC	AATGACCTGA	ATGAACAATT	GAAATGAAAA	2820
253	GAaaaaagta	TTGTTCCAAA	TTAAACGTTT	TAACCTTTTA	ATAGGTTTAT	ACAATAATTG	2880
254	ATATATGTTT	TCTGTATATG	TCTAATTTGT	TATCATCCAT	TTAGATATAG	ACAAAAAATA	2940
255	ATCTAAGAAC	TAAAACAAAT	GCTAATTTGA	AATGAAGGGA	GTATATATTG	GGATAATGTC	3000
256	GATGAGATCC	CTCGTAATAT	CACCGACATC	ACACGTGTCC	AGTTAATGTA	TCAGTGATAC	3060
257	GTGTATTAC	ATTTGTTGCG	CGTAGGCGTA	CCCAACAATT	TTGATCGACT	ATCAGAAAGT	3120
258	CAACGGAAGC	GAGTCGACCT	CGAGGGGGGG	CCCGGTACCC	AGCTTTTGT	CCCTTTAGTG	3180

09020716 "020998"